Journal of Quantitative Description: Digital Media 1(2021), 1–41 10.51685/jqd.2021.022

Problematic Content in Spanish-Language Comments on YouTube Videos About Venezuelan Refugees and Migrants

LUIS ENRIQUE AGUIRRE ZAPATA
Technische Universität Ilmenau, Germany

EMESE DOMAHIDI
Technische Universität Ilmenau, Germany

The present study identifies extensive content on YouTube relating to the recent Venezuelan refugee crisis, which has mostly affected neighboring countries, such as Peru and Ecuador. Furthermore, we use manual coding (n = 15,000) and computational text analysis (n = 217,028) to analyze user comments from 200 selected Spanish-language YouTube videos on the Venezuelan refugee crisis. Within this sample, we find an especially high number of problematic comments on videos about Venezuelan refugees and migrants, of which 32% were offensive comments and 20% were hateful comments. The most distinctive language characteristics reveal references to xenophobic, racist, and sexist content and show that offensive content and hate speech are not easily separated. While we identify many unique users (n = 95,915), only a small percentage (approximately 8%) of highly active users were responsible for approximately 40% of the identified problematic content. Highly active problematic users responded to each other more frequently than less active users and actively commented on multiple videos, indicating a network structure within our sample. The results of this study highlight the much-neglected discussion about Venezuelan refugees and migrants on YouTube. Furthermore, these results contribute to an enhanced understanding of online hate speech in the Latin

American context, which may lead to better and earlier hate speech detection and intervention.

Keywords: Problematic Content, Hate Speech, Venezuelan Refugees, Venezuelan Migrants, YouTube

The ongoing economic, political, and social situation in Venezuela has led to one of the largest emigration movements in the modern history of Latin America (Global Migration Data Portal, 2020; Welsh, 2018). The preference for intraregional migration is reflected in the enormous number of Venezuelan refugees and migrants in neighboring countries (e.g., Colombia: 1.3 million, Peru: 768,000, Chile: 228,000, Ecuador: 263,000, Brazil: 168,000, and Argentina: 130,000; UN High Commissioner for Refugees, 2019). To stem the migration wave, some Latin American countries (e.g., Chile, Ecuador, and Peru) changed their migration policies in 2019 and now require visas for Venezuelans (Sahhar, 2021). Not only are migration policies being adjusted in neighboring countries, but the public attitude toward migrants from Venezuela is growing increasingly harsh (Ramsey & Sánchez-Garzoli, 2018). Fear and aggression are common negative emotional responses to migrants (Erisen et al., 2020); these responses can increase with the amount of migration in a country (McLaren, 2003) and are often communicated through social media (Hrdina, 2016).

Although YouTube has been criticized for allowing its users to spread problematic content (Agarwal & Sureka, 2014), previous research on this topic has tended to focus on problematic Twitter content (Ripoll & Navas-Alemán, 2018; Rivero, 2019). User comments that engage with available content on a platform such as YouTube can bias audience perceptions (e.g., toward news; Lee & Jang, 2010), even before the actual content is accessed (Gearhart et al., 2020). While most studies have focused on problematic content in English (Malmasi & Zampieri, 2017), we argue that Spanish language and regional context are essential to understanding the discussion surrounding the current migratory crisis in Venezuela. Spanish is not only spoken by much of Latin America, but it is also

the native language of Venezuelans. Accordingly, Spanish-language problematic content (e.g., prejudiced commentary) is easily accessible to Venezuelans and thus may have negative consequences. For example, most instances of discriminatory behavior toward Venezuelans have been reported in Spanish-speaking countries in Latin America (International Organization for Migration [IOM], 2018e). While most prior studies of problematic online comments have focused on the content of these comments, initial research on user behavior has shown that networks exist between users who spread problematic content (Murthy & Sharma, 2019). Often, a large amount of problematic content comes from a small hateful user group (Evkoski et al., 2021; Mathew et al., 2019). Examining Spanish-language problematic content and user behavior on YouTube allowed us to shed light on the unique perspective of Spanish-speaking Latin America on the Venezuelan migration crisis and raise awareness of the impactful large-scale migratory movement of Venezuelans.

Prejudices and Discrimination Against Venezuelan Refugees and Migrants

Although the terms *refugee* and *migrant* are often used interchangeably, there are essential differences between them. The term *refugee* is legally defined as any person who, due to "well-founded fear of being persecuted" for a variety of reasons, "is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country" (UN High Commissioner for Refugees, 2010). A *migrant* is defined as someone who seeks to improve their quality of life by finding a place to work, a better education, and family reunion (Edwards, 2016).

Venezuelan refugees and migrants experience discrimination in various Latin American countries, mostly based on their nationality. The percentage of Venezuelan refugees or migrants who experience discrimination in Peru is 39%, in Ecuador 37%, in Chile 27%, in Argentina 7%, and in Brazil 6% (IOM, 2018a, 2018b, 2018c, 2018d, 2018e). These individuals also experience other forms of aggression, such as physical violence, sexual violence, and verbal aggression (IOM, 2018b, 2018d). These events can be traced back to prejudice, a hostile attitude toward certain individuals based on their belonging to

a certain group and based on negative characteristics associated with that group (Allport, 1979). Negative reactions to out-groups are often based on fear (Balleck, 2019). Migrants are often perceived as a threat by members of the majority population, who perceive them as competition for jobs and resources in times of economic disorder (Bobo, 1983) and immigration (Omi & Winant, 2015); thus, migrants face a variety of prejudices. The group-focused enmity (GFE) model depicts various types of prejudice as "a generalized devaluation of out-groups" (Zick et al., 2008, p. 367) and the denigration of these groups due to the characteristics assigned to them. Out-groups may persist across cultures (e.g., gender and age) or may evolve outside of a culture-specific or time-specific situation (e.g., migrants; Zick et al., 2008). Venezuelan refugees and migrants can be considered an out-group as the outcome of events occurring at a particular time (i.e., a country's current internal crisis). The GFE criteria that are relevant to the context of Venezuelan migrants (racism, xenophobia, sexism, and homophobia) due to situational, cultural, and country-specific factors will be outlined below, with a special focus on Latin American history and the current situation.

"[R]ace, migrant status, ethnicity, religion or belief, color, and other characteristics" can make a person the target of racists (Ghanea, 2012, p. 6). Historically, racialized language was employed in Latin America to denigrate citizens with African American ancestry, who were depicted "as inherently criminal, intellectually inferior, overly sexual, and animalistic" (Hernández, 2011, p. 816). Additionally, colonial racism (part of the Spanish colonial epoch's legacy) seeks to affirm the superiority of the conquerors' race and propagates the "cleaning of the blood" of the conquered group by mixing with a white conqueror to "whiten" their skin color (Manrique, 1999). Racialized language describes certain groups as *mestizos* (a mix of white European and indigenous), *mulattos* (white European and black), and *sambos* (black and indigenous). Similarly, *indio* and *cholo* refer to a person with an indigenous background or an indigenous peasant who immigrated to the city. Today, these traces of dominant ideologies are still present as parts of the racist narrative in Latin America (Rodríguez García, 2011).

Xenophobia refers to the rejection of what is perceived as culturally foreign (Crush & Ramachandran, 2010) and the hostile attitudes directed toward non-natives in each

population (UN Educational, Scientific and Cultural Organization [UNESCO], 2017). Typically, xenophobia is triggered when there is an influx of migrants within a community (European Monitoring Centre on Racism and Xenophobia, 1999). Latin American countries share similar ideological and cultural parameters (Rocha, 2018). However, the Venezuelan migratory wave has led to manifestations of xenophobia in this region, indicating detachment from the crisis, those affected by migration, and the Latin American identity. Xenophobia can be linked to the emergence of new relationships with neighboring countries, a lack of collective regional identity, and the denial of the out-group's right to form its own identity (Rocha, 2018). An example of a lack of regional identity is the minor role played by Latin American countries and regional organizations in facilitating a solution to the Venezuelan crisis. Ramon-Berjano (2011) mentioned that Latin American integration efforts have failed due to disparities in economic development between the countries, disputes caused by unsuccessful integration schemes and the inability to learn from these past mistakes, a lack of political commitment, trade disputes (e.g., a customs code agreement and several segmented trade agreements), and a lack of coordination between the countries.

One expression of sexism is sexist hate speech (The Council of Europe, 2016), which refers to the reduction of women to a sexual dimension or disdain for women based on their gender. Sexist hate speech can be presented online or offline through body shaming, revenge porn, violent and sexual death threats, or offensive comments about someone's appearance to ridicule and humiliate them (The Council of Europe, 2016; Ford, 2018). In Latin America, traditional gender roles for men and women are still widely prevalent. A particularly prevalent gender role for women is *marianismo*, which describes the role of a woman as that of one who is submissive, subordinate, and responsible for the family's well-being and spiritual growth (Nuñez et al., 2016). Moreover, discrimination, violence against women, and general inequality are prevalent in Latin America (Merkin, 2012).

Hate speech toward the LGBTI community is an expression of hatred, prejudice, and intolerance based on sexual orientation (Cowan et al., 2005). For example, LGBTI groups generally experience relatively high levels of negativity, unemployment, and anti-

gay violence (Diplacido, 1998). It is a common practice to delineate a homosexual person's behavior from what is perceived as "normal" heterosexual behavior (Plummer, 2001). The use of derogatory homophobic slurs, such as *faggot* (in Spanish, *maricón*), by heterosexual men has been shown to affirm their masculinity and self-identity (Carnaghi et al., 2011). The term *fag* is broadly used by homosexual and heterosexual men as an insult for men in general, regardless of their sexual orientation (Brown & Alderson, 2010). The same term is used in some Spanish-speaking countries—such as Ecuador, Peru, and Colombia—to insult heterosexual men or attack their manhood. Similarly, certain expressions are used to refer to and insult lesbians (Baére et al., 2015).

Other criteria from the GFE model (e.g., ageism, religion, veteran status, and disability) are not relevant to the present topic. Venezuelan immigrants in Latin America tend to be young (Mittelstadt, 2020), with an average age of 28 years (Saa et al., 2020). Therefore, ageism was excluded from this study. As in most Latin American countries, a Christian religious affiliation represents 88% of the entire region of Venezuela (Pew Research Center, 2014). Furthermore, all wars (not conflicts) in which Latin American countries were involved took place before the 20th century. Thus, veteran status does not play a role in this region. While little trustworthy data is available on the prevalence of disability in Venezuela (Dudzik et al., 2002; World Health Organization & World Bank, 2011), a comparison of Latin American countries showed that Venezuela's disability rates fall in the middle compared to those of other countries in the region (Economic Commission for Latin America and the Caribbean [ECLAC], 2013). Additionally, it can be assumed that mainly non-disabled Venezuelans have migrated to neighboring countries.

Characteristics of Problematic Content in User Comments on YouTube

In the context of user comments on the Internet, *hate speech* refers to problematic content that is used to dehumanize and diminish a group or one of its members by clearly creating a distinction between "them" and "us" (Gagliardone et al., 2015). More specifically, Álvarez-Benjumea and Winter (2018) identified six indicators of online hate: "1) contains negative stereotypes, 2) uses racial slurs, 3) contains words that are insulting,

belittling, or diminishing, 4) calls for violence, threat, or discrimination, 5) uses sexual slurs, and 6) sexual orientation/gender used to ridicule or stigmatize" (p. 8). In a nutshell, online hate speech is a form of expression that seeks to use online platforms to incite, promote, or spread hate based on intolerance, prejudices, intimidation, victimization, disapproval, discrimination, intimidation, violence, insults, degradation, humiliation, dehumanization, and diminishment toward a minority group (out-group) or a group member (individual) on the basis of characteristics such as gender (sexism), race (racism), age (ageism), religion, disability, veteran status, and sexual orientation (homophobia).

Hate speech is not the only type of problematic content on social media. On the contrary, YouTube users may engage in "a multitude of forms of hostile expression" (Murthy & Sharma, 2019), such as offensive statements or insults. (Schultes et al., 2013). Offensive content may contain strong or mild insults, slurs, or name-calling directed toward individuals without explicit reference to a particular group. This type of language can be uncivil, impolite, vulgar, aggressive, violent, or extremely rude. Additionally, this content may be used to denigrate someone and may contain negative adjectives (e.g., *dumb, ignorant,* and *criminal*) or insulting opinions. The main factor that differentiates offensive content from hate speech is that offensive content exhibits no explicit group-based discrimination of individuals. However, social media users often post short, grammatically incorrect messages. Therefore, the reason for which a particular message does not explicitly reference a group is often unclear. Furthermore, researchers tend to examine individual user comments and rarely analyze any references to previous messages. Therefore, the distinction between group-related hate speech and insults directed toward specific individuals is often difficult to determine, as both contain problematic content with similar characteristics.

Today, YouTube is the second-most-viewed website in the world (*Top sites*, 2019) and is widely regarded as a platform that is especially prone to individual users or groups promoting hate speech and offensive content (Agarwal & Sureka, 2014). User comments are a "computer-mediated, public, and interpersonal form of communication, which is published in connection with online content" (Schindler & Domahidi, 2021, p. 5). User comments that engage with the available content on a platform may impact audience

perceptions regarding this content, even before the actual content is accessed (Gearhart et al., 2020). Reading user comments on YouTube is related to information-seeking motives (Khan, 2017), and reading others' comments may affect users' personal opinions on the topic under discussion (Lee & Jang, 2010). Accordingly, our first research question aimed to reveal the extent to which Spanish-language user comments on YouTube videos about Venezuelan refugees and migrants contain hate speech and offensive content.

> *RQ1: To what extent do Spanish-language comments on YouTube videos about Venezuelan refugees and migrants contain hate speech and offensive content?*

Hateful user comments have been found on YouTube videos across various domains. For example, female YouTubers receive more hostile user comments than their male colleagues and are often victims of sexism (Döring & Mohseni, 2020). Furthermore, many YouTube comments have labeled Syrian immigrants as potential threats, traitors, and a source of financial hardship for the citizens of their host countries (Aslan, 2017). In Turkey and Poland, Sayimer and Derman (2017) used YouTube video comments to show how easily hate speech is transmitted and how fear of and violence toward migrants is incited. Racist, hateful user comments have even been identified in the context of counterspeech videos (Ernst et al., 2017). Negative reactions to out-groups (e.g., hate speech) are often based in fear and prejudice. As mentioned previously, the GFE model depicts various types of prejudice as "a generalized devaluation of out-groups" (Zick et al., 2008, p. 367) and the denigration of these groups due to the characteristics assigned to them. Venezuelan refugees and migrants can be considered an out-group that may experience exclusion based on racism, xenophobia, sexism, and homophobia. Our second research question aimed to qualify the insights derived from RQ1 by revealing the types of hateful comments that were present in our sample.

> *RQ2: Which types of hate speech (racism, xenophobia, sexism, or homophobia) are contained in Spanish-language user comments on YouTube videos about Venezuelan refugees and migrants?*

Within the unique political, economic, and cultural context of Latin America (Ramon-Berjano, 2011), the Venezuelan crisis is one of the largest migratory crises in modern history (Global Migration Data Portal, 2020; Welsh, 2018). Problematic YouTube content may be connected to both language-specific and context-specific factors (Garten et al., 2019). Therefore, inquiry into the language characteristics employed in this content is essential to understanding the discussion surrounding the Venezuelan crisis on YouTube. For example, word frequency can be revealed in the text classification process and employed in exploratory data analysis to provide a better understanding of the given context. In describing 1) the most frequent terms (i.e., words appearing several times in diferent comments), 2) frequent unique terms (i.e., words appearing in only one category) and 3) the most frequent bigrams (i.e., two consecutive words appearing several times in different comments), we acknowledge that language reflects a nation's social experiences, life, and culture (Geng, 2010) and that each language has its own unique means of expression (Lo et al., 2016). Word contexts are important in understanding hate speech. Stand-alone words may not be indicative of hate speech, but when combined, they can generate hate speech (Laaksonen et al., 2020). Our third research question allowed us to highlight the unique perspective of Spanish-speaking Latin America on the Venezuelan migration crisis and also allowed us to advance research on hate speech in this context.

> *RQ3: What are the most important language characteristics (i.e., the most frequent terms and bigrams) of hate speech and offensive content about Venezuelan refugees and migrants?*

Online hate speech and offensive comments tend to be more spontaneous than their in-person counterparts because users tend to post "instant responses, gut reactions, unconsidered judgments, off-the-cuff remarks, unfiltered commentary, and first thoughts" (Brown, 2018, p. 304). Anonymity, which is facilitated and thus common in an online context, may enhance a user's willingness to post hateful or offensive comments (Brown, 2018). Research on hate speech has mostly focused on the language-based analysis of hate

speech in user comments. However, initial research on user behavior in this context has revealed usage patterns that could help in understanding the dissemination of problematic content in a broader context. Studies (Matamoros-Fernández, 2017; Murthy & Sharma, 2019) have shown that YouTube serves as a social network for platform-based racism, wherein networked groups spread hate online. Users who make hostile comments connect with each other through various YouTube videos and post high numbers of hostile responses (Murthy & Sharma, 2019). Other studies have outlined that a large amount of problematic content on social media comes from a small hateful user group (for Twitter, see Evkoski et al., 2021; Mathew et al., 2019). Uncovering user patterns can not only help in understanding specific cases but can also improve hate speech detection and intervention. Accordingly, by addressing our fourth research question, we aimed to analyze the network structure of the distribution of problematic content through user comments on YouTube videos about Venezuelan refugees and migrants.

> *RQ4: Which users distribute problematic content, and how interconnected are they?*

Methods

Sample

To investigate our research questions, we conducted a case study of user comments from 200 videos on the current migratory crisis in Venezuela. First, we defined keywords for a search of relevant YouTube videos based on previous literature covering the terms *migrants* and *refugees* (UN High Commissioner for Refugees, 2019) and the current migratory crises in Venezuela (Baldwin, 2017). We aimed to cover the nationality of the migrants and refugees (*Venezuelan*), their country of origin (*Venezuela*), their migratory status (*refugee* or *migrant*), their place of destination (*in* or *to*), and migratory wave terminology (*exodus* or *diaspora*). Thus, we combined these five elements into the following eight search keywords: *refugiados Venezolanos* [Venezuelan refugees],

refugiados Venezuela [refugees Venezuela], *migrantes Venezolanos* [Venezuelan migrants], *migrantes Venezuela* [migrants Venezuela], *Venezolanos en* [Venezuelans in], *Venezolanos a* [Venezuelans to], é*xodo Venezolano* [Venezuelan exodus], and *diáspora Venezolana* [Venezuelan diaspora].

Next, we conducted a qualitative content analysis of the obtained videos and manually selected videos based on four inclusion criteria. 1) The most-viewed videos were selected based on their view counts. 2) As hate speech is language-dependent, the selected videos had to be in Spanish, the official language of most Latin American countries and the native language of most Venezuelans. 3) There was no restriction on the type of channel or video producer (e.g., vloggers, TV channels, or newspapers). The topic of each selected video had to cover the Venezuelan migratory crisis or be related to groups involved in the migratory crisis, such as Venezuelan refugees, Venezuelan migrants, or Venezuelans in other countries. Videos covering the Venezuelan humanitarian crisis, the Venezuelan political crisis, or the economic crisis (either in general or in Venezuela) were excluded if they did not mention the migratory perspective. Personal stories regarding the experiences of Venezuelans migrating or moving to different countries were included in the sample. Videos of Venezuelans making economic or political comparisons between Venezuela and other countries were excluded. 4) Videos that did not receive any comments or had disabled comment sections were excluded.

As a result of this qualitative content analysis, the final sample (200 videos) consisted of the 25 most-viewed videos on YouTube for each of the eight keywords. The selected videos consisted of vlogs, personal videos, interviews, testimonials, TV news, and newspaper videos from January 2015 to January 2019. A list of unique video identifiers was created, and then all the comments on each video were scraped using Python version 3.7.2. Altogether, 235,251 comments made by 101,481 unique users from January 2015 to February 2019 were scraped from all 200 videos. These comments consisted of 125,474 parent comments (first level) and 109,777 child comments (replies). These data were collected on February 5, 2019. All data, materials, and analysis scripts have been anonymized and are available in the supplementary online materials on OSF.[1]

[1] https://osf.io/8v5w2/

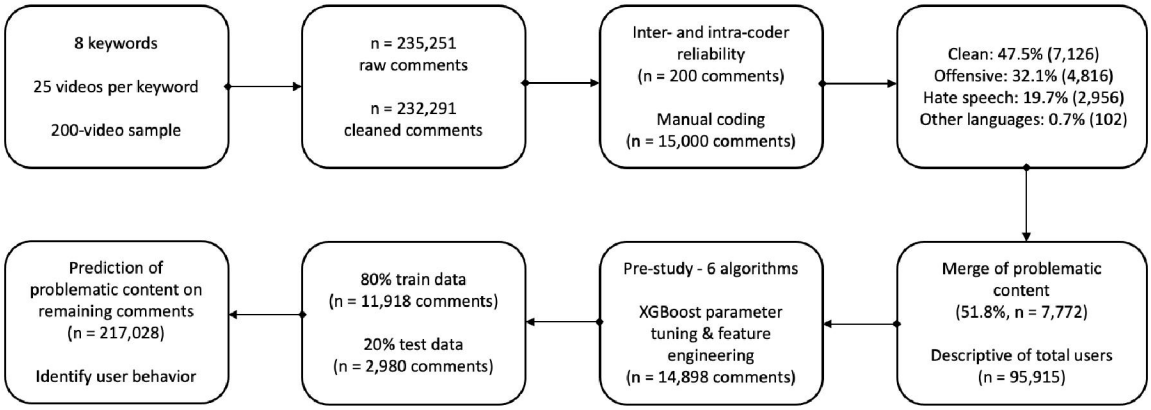

Figure 1: Summary of the data collection, coding, and prediction processes of the current study.

Data Cleaning

A standard data cleaning procedure was applied to the comments. First, we identified and removed URLs via the R package "qdapRegex" (Rinker, 2017). Comments that contained only a link and no text (n = 522) were removed from the data frame. Second, all ideograms (e.g., emojis) were identified via a list provided by Unicode version 12.0 (Unicode, 2019), and emoticons were identified via a list provided by the package "lexicon" (Rinker, 2018). All of these were removed from the text via the R package "stringi" (Gagolewski, 2019). Third, grammatically incorrect patterns—such as repeated characters of the same letter, repeated punctuation, missing spaces between words, and incorrect punctuation—were corrected in the comments via the R package "grep" (R Core Team, 2018). In the fourth step, most non-Spanish comments were deleted from the data frame. English comments were detected via the R package "cld3" (Ooms, 2018) and removed from the data frame (0.18% of all comments, n = 429). However, language detection does not perform well for non-English comments. Portuguese comments were identified based on special characters typical of Portuguese (e.g., "ã") using the package

"stringr" (Wickham, 2019) and then removed (0.31% of all comments, n = 733). Sixth, special symbols between one and three characters were removed from the data, as they were deemed irrelevant to the classification and manual coding phase. The cleaning process produced a final sample of 232,291 cleaned comments.

Analysis

On social media, the detection of problematic content poses various challenges—most importantly, the question of reliability (the quality of detection) and scalability (the capacity to process large amount of data). In previous studies, problematic social media content has been classified manually with human annotators and computational text analysis, specifically through machine learning algorithms (Kwok & Wang, 2013). A combination of these two methods is a particularly promising way of addressing both challenges and was thus used in the present study.

Manual Coding. A codebook (see online materials on OSF) helped structure the manual coding phase by providing examples and definitions to accurately analyze the data (DeCuir-Gunby et al., 2011). This codebook consisted of four blocks. The first block helped to exclude non-Spanish comments based on the approach of Cieliebak et al. (2017). The second block was dedicated to identifying the types of comments present on YouTube based on studies by Spertus (1997), Burnap and Williams (2016), and Gambäck and Sikdar (2017); these types included hate speech, offensive non-hate speech, and clean (i.e., non-offensive) comments. The third block was dedicated to identifying the targets of hate speech, such as racism, xenophobia, sexism, and homophobia (Zick et al., 2008). The fourth block was dedicated to formal categories, such as ID labels to identify comments.
An intercoder reliability test was performed between two coders on 100 YouTube comments via the R package "irr" (Gamer et al., 2019). The results of this test were satisfactory, with 92–100% agreement on the hate speech categories (racism, xenophobia, sexism, and homophobia) and kappa scores of 0.8–1 between the two coders. In terms of intracoder reliability (n = 100), the first coder reached 97–100% agreement on the hate

speech categories and kappa scores of 0.8–1. Next, a random sample of 15,000 comments was coded by both coders, who identified the variables of hate speech, offensive non-hate speech, clean (i.e., non-offensive) speech, and non-Spanish languages in the text. The manually detected hate speech comments from the previous phase (n = 2,956) were then coded according to the hate speech categories (sexist, racist, xenophobic, and homophobic comments) by one coder.

Computational Text Analysis. The computational text analysis began with common data preparation tasks (Ikonomakis et al., 2005)—such as removing stop words, punctuation, numbers, and excess empty spaces as well as stemming and lowercasing words—which were carried out via "quanteda" (Benoit et al., 2018) and "tm" (Feinerer & Hornik, 2008). The comments that were manually coded as *hate speech* and *offensive comments* were combined in the automatic text analysis stage due to their similar linguistic characteristics (see Figure 5) and their lack of explicit references to specific groups. We divided the manually coded data into 80% training data and 20% test data for model building and validation. Then, a corpus and a term frequency matrix were created to test the algorithm hyperparameters. The final vocabulary was reduced through rarely occurring words; thus, we achieved a term frequency matrix with 99% sparsity.

Extreme Gradient Boosting (XGBoost) has been found to perform accurately and efficiently in text classification tasks (Chen & Guestrin, 2016). Nevertheless, we conducted a pre-study to determine whether XGBoost could perform better than other algorithms with the provided data. In this pre-study, we ran five additional models based on five algorithms (support vector machine, random forest, LogitBoost, neural networks, and naive Bayes) using the data that was manually coded as hate speech. We found that XGBoost performed better in terms of both precision (0.50) and the F1 score (0.60) compared to the other algorithms (precision < 0.43; F1 < 0.57). Thus, we used XGBoost in the present study, and parameter tuning was performed via the package caret (Kuhn, 2019) and the five-step model of Pelkoja (2018). *Features* is a common expression that refers to the independent variables that are used to model an outcome (e.g., hate speech). The aim of feature selection is to reduce dimensionality and irrelevance in the features that will be inputted to a model;

thus, identity characteristics are useful in increasing data generalization (Ikonomakis et al., 2005). Various features were tested in the present study, including increased sparsity of the term document matrix (Feinerer, 2020), selective stop word exclusion/inclusion, stemming and lemmatization (Korenius et al., 2004; Torres-Moreno, 2012), tf vs. tf-idf weighting (Ramos, 2003), terms vs. bigrams (Burnap & Williams, 2015), part-of-speech tagging, sentiment analysis (Schmidt & Wiegand, 2017), and various combinations of these features. Due to the imbalanced data structure (Ganganwar, 2012), upsampling and downsampling techniques were applied to the train data. Altogether, 27 models were trained and compared in terms of precision, recall, accuracy, kappa scores, and F1 scores, which are all standard methods of determining the effectiveness of a classification (Ikonomakis et al., 2005). Our final model for data classification exhibited a kappa score of 0.56, a precision of 0.83, a recall of 0.72, and an F1 score of 0.77. This model classified the combination of hate speech, and offensive comments, by using terms with customized stop words in Spanish, stemming, tf-idf weighting, and no sampling (see online materials on OSF).

The trained model with the XGBoost algorithm was run on the remaining cleaned data sample (217,028 comments) to automatically predict problematic content. For an overview of the data gathering and analysis procedure, see Figure 1.

Results

Here, we provide an overview of the manual coding results and describe the prevalence of problematic content (including hate speech and offensive content) in user comments on YouTube videos about Venezuelan refugees and migrants. We also discuss the most important characteristics of hate speech and offensive content and describe the types of hate speech found in the analyzed comments. Lastly, we highlight the results of the automatic text analysis, with a special focus on user behavior.

Prevalence and Most Important Characteristics of Problematic Content

The present analysis relied on a random sample of 15,000 manually coded comments, which represented the population of comments in our data set.

Prevalence of Problematic Content. Of all the manually coded comments, 47.5% (n = 7,126) were clean comments that included information, facts, or mere opinions without offensive or hateful content. Offensive comments represented 32.1% (n = 4,816) of the sample and contained insults, offenses, or impolite words but did not reference any particular groups. Moreover, 19.7% (n = 2,956) of the data consisted of hateful comments, which contained offenses, insults, and derogatory terms and explicitly referenced particular groups. Manually coded comments that contained non-Spanish languages (0.7% of all manually coded comments, n = 102) were excluded. Accordingly, 51.8% of all comments in our sample contained some form of problematic content. Comments containing hate speech received higher numbers of likes (M = 3.2, SD = 25.5) than offensive (M = 2.3, SD = 15.7) or clean comments (M = 2.3, SD = 22.6).

Hate Speech. In 19.7% of the comments in our sample, we found various types of hate speech. Xenophobia predominated relative to other types of hate speech, representing 83% (2,461) of the entire sample (n = 2,956) of hate speech comments. Examples[2] of xenophobic comments in our sample included the following: “You are already too many Venezuelan pigs! Don’t come here anymore. You only come to steal and increase crime! Stay in your country!” In terms of prevalence, xenophobic comments were followed by racist comments (16%, n = 478; e.g., “Don’t come to Peru, you black son of a bitch” or “Venezuelan monkeys why don’t you emigrate to your mother country Congo?”) and sexist comments (15%, n = 432; e.g., “You must be one hell of a whore... like every Venezuelan

[2] The Spanish-language discussion of the Venezuelan migration crisis on YouTube includes a large amount of problematic content. In order to describe this discussion, we provide examples of such problematic content (indicated by the quotation marks). Unfortunately, these examples contain content that may be offensive or objectionable. Nevertheless, we believe it is important to understand the discussion in order to identify harmful content in the future and mitigate its impact. Accordingly, we report these examples and make only the worst examples unidentifiable.

woman, loves cock"). Only a few cases of homophobia were detected (4%, n = 125; e.g., "How do you know I'm not a man, did you suck my dick? And regarding the spelling that does not offend me faggot asshole").

In the next step, we aimed to compare the language characteristics of hate speech and offensive content to reveal the nature of problematic comments (see Figure 2). We found that similar words (e.g., *Venezuelan*, *country*, and *Peru*) appeared frequently in hate speech, offensive content, and clean content. Thus, to distinguish hate speech from offensive content, we will discuss and put in context the most frequent unique terms for each category. The top 20 unique terms for the hate speech comments (i.e., words that did not appear in offensive or clean comments) included words related to others' sexual orientations/health, such as "*faggot", "AIDS sufferer", "transgender", "sissy",* and "*gay"*. Another common pattern was shaming other users for their affinity with Venezuelans by using the term "*veneco lover"*. Racist slurs directed toward black Venezuelans (e.g., "*ape"* and *"n****r"*[3]) and indigenous people pointed toward classic racist patterns described for the Latin American context (Hernández, 2011; Manrique, 1999). At least ten terms were used to attack or diminish the Venezuelan nationality or country, and two terms ("*troublemaker"* and *"dog eater"*) were used to describe the behavior of a group of Venezuelans. Despite the high frequency of the word "*faggot"* (homophobia) as a unique term, there were more terms with xenophobic and/or racist connotations of hate speech in the sample. As shown in Figure 2, the present study considered many offensive words, most of which were not at all frequent in our sample. This means that hate speech can be expressed in many ways and is represented differently in different user comments.

[3] Word partially censored due to its extremely offensive racial connotation

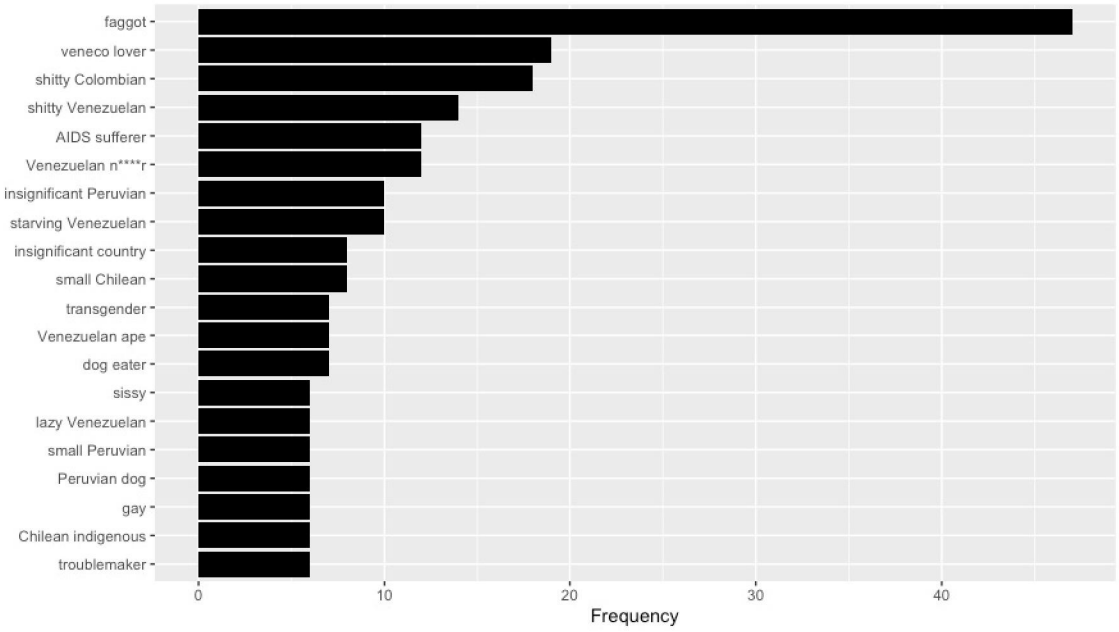

Figure 2: The top 20 unique terms for comments manually coded as hate speech. *Unique* denotes words that appeared only in hate speech comments. Depicted are the translated terms in English. Most of the terms were composed of two words (e.g., *negrozolano*, a combination of the words *negro* [n****r] and *venezolano* [Venezuelan]). Moreover, some of the identified words had no literal English translations (e.g., *paisucho*,

To provide more context and better understand hate speech narratives, we highlighted the most frequent bigrams (i.e., sequences of two adjacent words) in our sample. Xenophobic (anti-migration) linguistic cues in the bigrams' word network (see online materials on OSF) included "*stop entering"*, "*close border"*, "*closed border"*, "*take away jobs"*, "*Venezuelan steal"* (or the extended version, "*Venezuelan only come steal"*), "*stay country"*, "*shitty country"*, and "*Venezuelan shitty"*. Certain generalizations were also found in the comments, such as "*Venezuelan are"*, "*all Venezuelan",* "*rest of the countries"*, and "*all the country"*. Additionally, some of the bigrams contained racist narratives (e.g., "*improve race"*, implying that one race is inferior). Others referred to the behavior of migrants (e.g., "*talk bad"*, "*ungrateful"*, or "*need to be more"*, and "*Venezuelan mor"e*) or expressed a desire to reclaim their own country and reject migrants

(e.g., "*fight for the/your country"*, *"stop enter"*, *"close border"*, and "*nobody wants"*). Insults, which often carried sexist connotations (e.g., "*son of a bitch"*), were less frequent than xenophobic and racist narratives.

Offensive Content. The top 20 unique terms for offensive content (i.e., words that did not appear in hate speech or clean comments) were identified and depicted in Figure 3. These terms included impolite expressions directed toward individuals (e.g., "*educate yourself"*, *"balls"*, and "*sheep"*), explicit insults directed toward individuals (e.g., "*idiot"*, *"bloody"*, "*abnormal"*, and *"excrement"*), and terms that urged users to leave their countries or disappear (e.g., "*get out"*, *"contraceptives"*, *"go around"*, and "*return"*). Additionally, we found some terms related to politics (e.g., "*candidate"*, *"Pinochet"*, *"dome"*, and *"mercenary"*), which pointed at least partly to some kind of political discussion surrounding the present topic. Comments containing political terms included those that blamed citizens for the elected president ("They are to blame for electing that donkey as president"); expressed hatred for politicians ("Damn Maduro and Diosdado Cabello, they are Satan, they are assassins"); and expressed hatred for political orientations such as the Left ("Another country that is in crisis because of the Left"), socialism ("I knew these socialist parasites had to go out and bray"), communism ("Shut up you fucking communist, continue to protect the political ideology that destroys homes, families and entire countries"), and dictatorship ("Dirty, discriminatory pig. When Chileans were escaping from Pinochet's dictatorship, the world welcomed them with open arms and helped them in every way possible"). Although the number of words with political connotations was higher than the number of insults, the word "*idiot"* had the highest frequency.

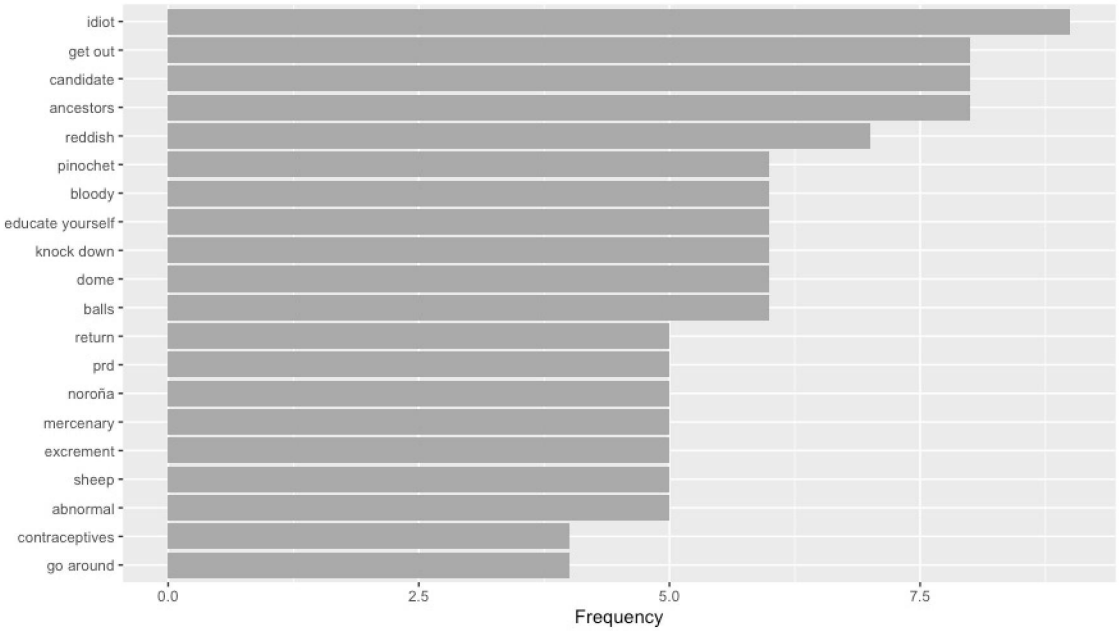

Figure 3: The top 20 unique terms for comments manually coded as offensive content. *Unique* denotes words that appeared only in offensive comments. Depicted are the translated terms in English.

Problematic Content. Next, we identified the 20 most common terms in both categories to compare the linguistic patterns of hate speech and offensive comments (see Figure 4). Twelve terms were present in both categories, including insults and terms that fell into categories of hate speech (e.g., racism and xenophobia). Insulting terms (e.g., *disgusting*), some of which carried sexist connotations (e.g., "*son of a bitch"* and "*motherfucker"*), appeared frequently in both categories. Terms with racist (e.g., "*ape"*), xenophobic (e.g., "*lazy ass"*), and sexual (e.g., "*dick sucker"*) connotations were also common in both categories. It is clear that hate speech and offensive comments overlap in terms of their most frequently used words, and both categories represent problematic content that may be highly insulting to recipients.

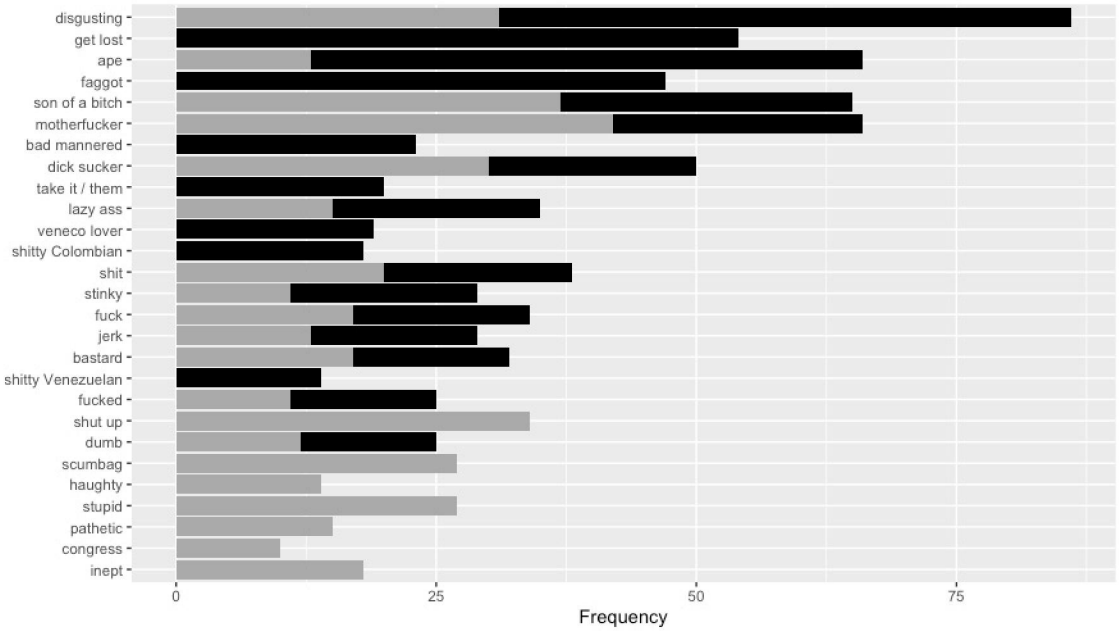

Figure 4: The top 20 terms for comments manually coded as hate speech and offensive content. Note: Words that appeared in clean comments were excluded. The color black represents hate speech, and the color grey represents offensive content.

Lastly, we looked at the bigrams of problematic content to reveal narratives in the bigrams' word network (see online materials on OSF). The word "*shit"* was broadly connected to several words with racist (e.g., "*n****r"* and *"indigenous"*) and xenophobic (e.g., "*country"*, *"veneco"*, *"Venezuelan"*, and *"Colombian"*) connotations in the bigrams' word network. Xenophobia was also found in references to the perceived behavior of Venezuelans (e.g., "*arrogant"*, *"coward"*, and *"bitch"*) as well as references to behavior without any targets (e.g., "*screwing Peru"*, "*screwing country"*, "*get out country"*, *"go to Peru"*, *"fucking arrive"*, *"want for free"*, and *"talk bullshit"*). Sexist insults made up a large portion of the terms found in problematic content, often targeting the word "*mother"* with words such as "*fuck"*, *"pussy"*, *"bitch"*, *"c**t"*[4], and "*son of a*

[4] Word partially censored due to its extremely offensive sexist connotation

whore". General insults included *so stupid*, *so ignorant*, and violent bigrams (e.g., "*kill Maduro"* [the president of Venezuela]).

User Behavior. As outlined above, offensive comments and hate speech share various characteristics; thus, we combined both categories for automatic text analysis to achieve a good model fit. Then, we used the developed supervised machine learning model on our remaining data set of 217,028 comments to automatically predict problematic content and identify user behavior. Altogether, 44% (n = 96,194) of the automatically classified user comments in our sample contained problematic content, while 56% were classified as clean comments.

Overall, 95,915 unique users commented on the videos about Venezuelan refugees in our sample, and 47.6% (45,718) of these users shared problematic content. Parent (47,415) and child (48,779) comments held a nearly 1:1 ratio (49.3% : 50.7%), meaning that users who shared problematic content not only created new comments but also made an equal number of comments to reply to other users. To understand the behavior of these replies, we analyzed the intensity of commenting per problematic user. Each problematic user commented twice on average (M = 2.10, SD = 3.85); thus, the problematic users were more active than the clean users (M = 1.74, SD = 6.34) in our sample. Unique problematic users replying to comments from their peers represented 82.9% (27,256) of all problematic content replies, while 17.1% (5,609) was accounted for by unique clean users. However, we also found a pattern that is common on social media platforms, in that only a few problematic users were responsible for most of the problematic content.

We defined *active users* as every user who published twice as many comments as the average. Only 7.84% of all problematic users (3,585) made more than four comments (M = 10.3, SD = 10.5), making up 38.40% (36,942) of all problematic comments. This made them highly active users sharing hate speech and offensive content. In the next step, we looked at whether these highly active users focused on only one video or whether we could reveal patterns of networked hate between different videos. We found that only a small minority of highly active problematic users (680) commented on only one video, while highly active problematic users commented on 5.08 videos on average (SD = 4.37).

Of all the highly active problematic users, 67.45% (2,418) commented on more than two videos in our sample, with the maximum number of videos commented upon by a single user being 45 videos; this clearly indicates a networked structure in the problematic content (see Table 1).

Table 1. Overview of clean users, problematic users, and highly active problematic users in our sample.

	Clean users	Problematic users	Highly active problematic users
n	50,197	45,718	3,585
Problematic content	-	96,194	36,942
Mean comment count	1.74 (SD = 6.34)	2.10 (SD = 3.85)	10.3 (SD = 10.5)
Mean videos commented	1.1 (SD = 0.4)	1.78 (SD = 1.84)	5.08 (SD = 4.37)

Discussion

This study was the first to describe problematic content consisting of hate speech and offensive user comments on YouTube videos about Venezuelan migrants and refugees. Regarding *RQ1* (*To what extent do Spanish-language comments on YouTube videos about Venezuelan refugees and migrants contain hate speech and offensive content?*), similar to prior literature regarding Twitter (Fundamedios, 2018), we found that a high percentage (51.8%) of the comments in our sample contained some sort of problematic content. Of this problematic content, 32.1% consisted of offensive comments containing insults, offenses, or impolite words but made no explicit reference to any particular group. Moreover, 19.7% of the comments were hateful, containing offenses, insults, and derogatory terms directed toward minorities and making explicit reference to a particular group. Previous research has shown that problematic online content can have serious

offline consequences (Müller & Schwarz, 2017; Oksanen et al., 2014), particularly for vulnerable groups such as Venezuelan migrants in Latin America.

Regarding *RQ2* (*Which types of hate speech [racism, xenophobia, sexism, or homophobia] are contained in Spanish-language user comments on YouTube videos about Venezuelan refugees and migrants?*), xenophobia was highly predominant in the data, making up 83% of the entire sample of hate speech comments. Racism and sexism followed, making up 16% and 15% of the hate speech comments, respectively. Only a small percentage (4%) of the hate speech comments exhibited homophobia. The GFE model does not cover prejudice toward groups based on political orientation; thus, this was not the focus of our study. Nevertheless, our results showed that references to political orientation and politicians could be key to better understanding problematic content in the context of migration. Therefore, future studies should consider prejudice toward groups based on political orientation to improve the detection of problematic content.

Regarding *RQ3* (*What are the most important language characteristics [i.e., the most frequent terms and bigrams] of hate speech and offensive content about Venezuelan refugees and migrants?*), we found patterns of problematic content similar to those found in prior literature. For example, we revealed words that referred specifically to race, such as *indigenous shit*, *improve race* (Manrique, 1999), and animalistic references (e.g., *ape*; Hernández, 2011). Other characteristics (e.g., *arrogant*) targeted undesirable qualities or unwanted behaviors (Parekh, 2006) perceived in Venezuelans. Problematic content also attacked unidentified targets for their behavior or presumed qualities (e.g., "*disgusting people", "bad-mannered", "lazy", "stinky", "want for free"*, and "*ungrateful"*). Important as well were characteristics related to the exclusion (Parekh, 2006) of Venezuelans, exhibited in narratives such as "*get out of the country"* or *"close the border"*. Furthermore, prejudiced statements that migrants are involved in illegal activities and prostitution (e.g., "*only come to steal"*; ILO, IOM, & OHCHR, 2001) were also identified as frequent narratives. Insults were highly prevalent in the problematic content, especially the slur "*faggot",* which was broadly used (Carnaghi et al., 2011) to spread hate based on sexual orientation (Álvarez-Benjumea & Winter, 2018).

A detailed view of the narratives of hate speech revealed that the xenophobic narratives involved the two dimensions described by Radkiewicz (2003). The first dimension focused on national superiority and insulting the migrants' country with words such as "*shit country"*. The second dimension included hostile behavior against other cultures via words such as *Venezuelan shit*. Furthermore, particular expressions (e.g., "*need to be more"*, and "*Venezuelan more"*) denoted that certain actions and behaviors should be carried out by a particular group. Our in-depth investigation of the Venezuelan migration crisis opened up an interesting perspective on the GFE model. In particular, our findings revealed xenophobic and racial biases in the discussion surrounding Venezuelan refugees and migrants in Latin America. This was initially surprising because in Latin America, it is not only language that is shared but also culture, religion, and racial diversity. These findings bring complexity to an already established out-group model like GFE, in which clear differences (e.g., religious differences) between in-groups and out-groups play a fundamental role in explaining prejudice toward out-groups. These differences may be less clear for regional migratory movements. This finding calls for the careful adaptation of out-group models to the specific context at hand.

Our results showed that economic fears—for example, that *they* (presumably Venezuelan refugees and migrants) are "*miserable", "take jobs away"*, "*talk bad about the country"* (presumably their host country), or are "*ungrateful"*—may be decisive in the present context. This tendency to describe immigrants in terms of their low income was also pointed out by Olmos Alcaraz (2018). While our work provides valuable initial insights into an under-researched area, the present analysis remains entrenched in the bag-of-words approach (Manning & Schütze, 1999). We have described frequent words and bigrams, but the broader context in which they are used remains unclear to us. Recent developments in computational linguistics, such as word embeddings (Bolukbasi et al., 2016; Garg et al., 2018), may help in more accurately identifying the biases and stereotypes associated with each out-group in YouTube user comments (Kroon et al., 2020).

In the present work, we found that hate speech and offensive comments share many characteristics and are thus hard to separate. In a social media context, it often remains unclear why certain comments may not contain explicit references to any particular group,

the main characteristic separating hate speech from offensive content. Accordingly, we used computational text analysis to classify problematic content in our full sample of 217,028 comments. This step helped shed light on user behavior in the data set. Regarding *RQ4* (*Which users distribute problematic content, and how interconnected are they?*), it was revealed that 47.6% of all users in our data set shared problematic content. Additionally, these users were highly interconnected with each other and with various videos. As in previous work (Evkoski et al., 2021; Mathew et al., 2019), we found that only a small percentage (approximately 8%) of highly active users were responsible for approximately 40% of the problematic content, and these users actively commented on multiple videos. These findings clearly indicated a networked structure in the problematic content, a result already highlighted by previous studies (e.g., on racism; Murthy & Sharma, 2019). Our results also confirmed that user behavior, alongside language-based analysis, is helpful in understanding problematic user comments and developing intervention measures. In particular, future studies should focus on highly active users, who spread the majority of hate speech. It would be fruitful to analyze more information about these users (e.g., user profile characteristics) and employ further network measures (e.g., centrality), preferably through a full survey of videos on the present topic. On the other hand, including user behavior in the classification of hate speech could help in identifying and potentially removing problematic accounts. Future research may include other measures—such as the number of videos viewed by each user, the geographical locations of users, and public discourse on other social media platforms (e.g., Twitter or Facebook) and traditional media—to more intricately qualify problematic content and user behavior in user comments on YouTube videos about Venezuelan refugees and migrants.

Additionally, while it is crucial that scientists use language and contextual knowledge to understand and interpret certain expressions or words in manual content analysis, we found that there are more challenges involved in computational Spanish-language text analysis. Not only have most studies of problematic content focused on the English language (Malmasi & Zampieri, 2017), but so does natural language processing in general; consequently, most of the tools developed for this purpose (i.e., stop word lists and Part-of-speech [POS] tagging) are for English text. While there are some tools for the

Spanish language, they are certainly not optimal, especially in a social media context where the quality of the language is extremely poor and most comments contain grammatical mistakes, a lack of punctuation, typos, or slang. To progress in this area of research, more tools with a focus on the Spanish language and the Latin American context will be needed.

As outlined above, the present study focused on user comments written in Spanish, which is the official language of most Latin American countries, the maternal language of most Venezuelans, and the language in which most discrimination toward Venezuelans has been reported (IOM, 2018e). To understand the present topic in a broader context, future studies could include Brazilian Portuguese in their analyses of problematic YouTube content to take into account the largest country in the region, which also has a large number of Venezuelan immigrants. New insights into multilingual computational content analyses (e.g., Chan et al., 2020; Lind et al., 2019) may enhance this endeavor. We hope that the present study's initial insights into the specifics of problematic content in the Venezuelan context will be helpful in these tasks. Accordingly, another contribution of the present study is the publication of materials for coding and classification that will enable future research on the present topic.

The present study successfully reached a better understanding of problematic content narratives in Spanish-language user comments on YouTube videos about Venezuelan refugees and migrants. Given the magnitude of the Venezuelan migration crisis, the high prevalence and nature of problematic content in YouTube comments, and the potential negative impact of hate speech and offensive content, it is crucial to widen the perspective of communication science to include events integral to the Latin American context, which are often overlooked.

References

Agarwal, S., & Sureka, A. (2014). A focused crawler for mining hate and extremism promoting videos on YouTube. *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 294–296. https://doi.org/10.1145/2631775.2631776

Allport, G. W. (1979). *The nature of prejudice*. Addison-Wesley Pub. Co.

Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, *34*(3), 223–237. https://doi.org/10.1093/esr/jcy005

Aslan, A. (2017). Online hate discourse: A study on hatred speech directed against Syrian refugees on YouTube. *Journal of Media Critiques*, *3*(12), 227–256. https://doi.org/10.17349/jmc117413

Baére, F., Zanello, V., & Romero, A. C. (2015). Xingamentos entre homossexuais: Transgressão da heteronormatividade ou replicação dos valores de gênero? [Homosexual gossip: Transgression of heteronormativity or replication of gender values?]. *Revista Bioética*, *23*(3), 623–633. https://doi.org/10.1590/1983-80422015233099

Baldwin, L. (2017). The Venezuelan diaspora: A cerebral exodus. *Latin American Studies: Student Scholarship & Creative Works*, 14. https://digitalcommons.augustana.edu/cgi/viewcontent.cgi?article=1000&context=ltamstudent

Balleck, B. J. (2019). *Hate groups and extremist organizations in America: An encyclopedia*. ABC-CLIO, LLC.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bobo, L. (1983). Whites' opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology*, *45*(6), 1196–1210. https://doi.org/10.1037/0022-3514.45.6.1196

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to

computer programmer as woman is to homemaker? Debiasing word embeddings. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 29, 4356–4364. https://papers.nips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, *18*(3), 297–326. https://doi.org/10.1177/1468796817709846

Brown, T. L., & Alderson, K. G. (2010). Sexual identity and heterosexual male students' usage of homosexual insults: An exploratory study. *Canadian Journal of Human Sexuality*, *19*(1-2), 27 –42.

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making: Machine classification of cyber hate speech. *Policy & Internet*, *7*(2), 223–242. https://doi.org/10.1002/poi3.85

Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science*, *5*(1), 11. https://doi.org/10.1140/epjds/s13688-016-0072-6

Carnaghi, A., Maass, A., & Fasoli, F. (2011). Enhancing masculinity by slandering homosexuals: The role of homophobic epithets in heterosexual gender identity. *Personality and Social Psychology Bulletin*, *37*(12), 1655–1665. https://doi.org/10.1177/0146167211424167

Chan, C.-H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., van Atteveldt, W., & Althaus, S. L. (2020). Reproducible extraction of cross-lingual topics (rectr). *Communication Methods and Measures*, *14*(4), 285–305. https://doi.org/10.1080/19312458.2020.1812555

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings*

of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 785–794. https://doi.org/10.1145/2939672.2939785

Cieliebak, M., Deriu, J. M., Egger, D., & Uzdilli, F. (2017). A Twitter corpus and benchmark resources for German sentiment analysis. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 45–51. https://doi.org/10.18653/v1/w17-1106

Council of Europe. (2016). *Combating sexist hate speech* [Fact sheet]. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680651592

Cowan, G., Heiple, B., Marquez, C., Khatchadourian, D., & McNevin, M. (2005). Heterosexuals' attitudes toward hate crimes and hate speech against gays and lesbians: Old-fashioned and modern heterosexism. *Journal of Homosexuality*, *49*(2), 67–82. https://doi.org/10.1300/j082v49n02_04

Crush, J., & Ramachandran, S. (2010). Xenophobia, international migration and development. *Journal of Human Development and Capabilities*, *11*(2), 209–228. https://doi.org/10.1080/19452821003677327

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods*, *23*(2), 136–155. https://doi.org/10.1177/1525822x10388468

Diplacido, J. (1998). Minority stress among lesbians, gay men, and bisexuals: A consequence of heterosexism, homophobia, and stigmatization. In G. M. Herek (Ed.), *Stigma and sexual orientation: Understanding prejudice against lesbians, gay men, and bisexuals* (pp. 138–159). https://doi.org/10.4135/9781452243818.n7

Döring, N., & Mohseni, M. R. (2020). Gendered hate speech in YouTube and YouNow comments: Results of two content analyses. *Studies in Communication and*

Media, *9*(1), 62–88. https://doi.org/10.5771/2192-4007-2020-1-62

Dudzik, P., Elwan, A., & Metts, R. (2002). *Disability policies, statistics, and strategies in Latin America and the Caribbean: A review*. https://unipd-centrodirittiumani.it/public/docs/31863_statistics.pdf

Edwards, A. (2016, November 7). *UNHCR viewpoint: "Refugee" or "migrant" – Which is right?* UNHCR. https://www.unhcr.org/news/latest/2016/7/55df0e556/unhcr-viewpoint-refugee-migrant-right.html

Erisen, C., Vasilopoulou, S., & Kentmen-Cin, C. (2020). Emotional reactions to immigration and support for EU cooperation on immigration and terrorism. *Journal of European Public Policy*, *27*(6), 795–813. https://doi.org/10.1080/13501763.2019.1630470

Ernst, N., Engesser, S., Büchel, F., Blassnig, S., & Esser, F. (2017). Extreme parties and populism: An analysis of Facebook and Twitter across six countries. *Information, Communication & Society*, *20*(9), 1347–1364. https://doi.org/10.1080/1369118X.2017.1329333

European Monitoring Centre on Racism and Xenophobia. (1999). *Looking reality in the face: The situation regarding racism and xenophobia in the European community*. https://fra.europa.eu/sites/default/files/fra_uploads/1944-AR_1998_part2-en.pdf

Evkoski, B., Pelicon, A., Mozetic, I., Ljubesic, N., & Novak, P. K. (2021). Retweet communities reveal the main sources of hate speech. *ArXiv*, *abs/2105.14898*. https://arxiv.org/pdf/2105.14898.pdf

Feinerer, I. (2020). Introduction to the tm package text mining in R. https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

Feinerer, I., & Hornik, K. (2008). *tm: Text mining package* (R package version 0.7-6) [Computer software]. https://CRAN.R-project.org/package=tm

Ford, C. (2018). *Fight like a girl*. Oneworld Publications.

Fundamedios. (2018). *La migración Venezolana acapara la conversación en Twitter [Venezuelan migration monopolizes the conversation on Twitter]*. https://www.fundamedios.org.ec/wp-content/uploads/2018/09/ESPECIAL.pdf

Gagliardone, I., Gal, D., Alves, T., & Martínez, G. (2015). *Countering online hate speech*. Unesco Publishing.

Gagolewski, M. (2019). *R package stringi: Character string processing facilities*. www.gagolewski.com/software/stringi/

Gambäck, B., & Sikdar, U. K. (2017). Using convolutional neural networks to classify hate-speech. *Proceedings of the First Workshop on Abusive Language Online*, 85–90. https://doi.org/10.18653/v1/w17-3013

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement* (R package version 0.84.1) [Computer software]. https://CRAN.R-project.org/package=irr

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering*, 2*(4), 42–47.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Garten, J., Kennedy, B., Sagae, K., & Dehghani, M. (2019). Measuring the importance of context when modeling language comprehension. *Behavior Research Methods*, *51*(2), 480–492. https://doi.org/10.3758/s13428-019-01200-w

Gearhart, S., Moe, A., & Zhang, B. (2020). Hostile media bias on social media: Testing

the effect of user comments on perceptions of news bias and credibility. *Human Behavior and Emerging Technologies, 2*(2), 140–148. https://doi.org/10.1002/hbe2.185

Geng, X. (2010). Cultural differences influence on language. *Review of European Studies*, *2*(2), 219–222. https://doi.org/10.5539/res.v2n2p219

Ghanea, N. (2012). *The concept of racist hate speech and its evolution over time*. https://www.ohchr.org/Documents/HRBodies/CERD/Discussions/Racisthatespeech/NazilaGhanea.pdf

Global Migration Data Portal. (2020, December 4). *Migration data in South America*. https://migrationdataportal.org/regional-data-overview/migration-data-south-america

Hernández, T. K. (2011). Hate speech and the language of racism in Latin America: A lens for reconsidering global hate speech restrictions and legislation models. *University of Pennsylvania Journal of International Law*, *32*(3), 805-801. https://ir.lawnet.fordham.edu/faculty_scholarship/19

Hrdina, M. (2016). Identity, activism and hatred: Hate speech against migrants on Facebook in the Czech Republic in 2015. *Naše Společnost*, *1*(14), 38. https://doi.org/10.13060/1214438x.2016.1.14.265

Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS Transactions on Computers*, *4*(8), 966–974.

International Labour Office, International Organization for Migration, & Office of the United Nations High Commissioner for Human Rights. (2001). *International migration, racism, discrimination and xenophobia*. www.refworld.org/docid/49353b4d2.html

International Organization for Migration. (2018a). *DTM Argentina - Monitoramento do*

fluxo migratório Venezuelano [DTM Argentina - Migration flows from Venezuela]. https://displacement.iom.int/system/tdf/reports/DTM%20Argentina%20Ronda%201.pdf?file=1&type=node&id=4293

International Organization for Migration. (2018b). *DTM Brasil - Monitoramento do fluxo migratório Venezuelano [DTM Brazil - Migration flows from Venezuela]*. https://migration.iom.int/system/tdf/reports/MDH_OIM_DTM_Brasil_N1_0.pdf?file=1&type=node&id=3522

International Organization for Migration. (2018c). *DTM Chile - Monitoreo de flujo de población Venezolana. [DTM Chile - Migration flows from Venezuela]*. https://migration.iom.int/system/tdf/reports/DTM%20Chile%20Round%201%20FINAL.PDF?file=1&type=node&id=4295

International Organization for Migration. (2018d). *DTM Ecuador - Monitoreo de flujo de población Venezolana. [DTM Ecuador - Migration flows from Venezuela]*. http://oim.org.ec/pdf/DTM%20Ronda%202.pdf

International Organization for Migration. (2018e). *DTM Perú - Monitoreo de flujo de población Venezolana. [DTM Peru - Migration flows from Venezuela]*. https://migration.iom.int/system/tdf/reports/DTM_R4_OIMPERU_VFF.pdf?file=1&type=node&id=4890

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior, 66*, 236–247. https://doi.org/10.1016/j.chb.2016.09.024

Korenius, T., Laurikkala, J., Järvelin, K., & Juhola, M. (2004). Stemming and lemmatization in the clustering of Finnish text documents. *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management - CIKM '04*, 625. https://doi.org/10.1145/1031171.1031285

Kroon, A. C., Trilling, D., & Raats, T. (2020). Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism & Mass Communication Quarterly, 98*(2), 451–477. https://doi.org/10.1177/1077699020932304

Kuhn, M. (2019). *caret: Classification and regression training* (R package version 6.0-84) [Computer software]. https://CRAN.R-project.org/package=caret

Kwok, I., & Wang, Y. (2013). Locate the hate: Detecting tweets against blacks. *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, 1621–1622*.

Laaksonen, S.-M., Haapoja, J., Kinnunen, T., Nelimarkka, M., & Pöyhtäri, R. (2020). The datafication of hate: Expectations and challenges in automated hate speech monitoring. *Frontiers in Big Data*, *3*, 3. https://doi.org/10.3389/fdata.2020.00003

Lee, E.-J., & Jang, Y. J. (2010). What do others' reactions to news on Internet portal sites tell us? Effects of presentation format and readers' need for cognition on reality perception. *Communication Research*, *37*(6), 825–846. https://doi.org/10.1177/0093650210376189

Lind, F., Eberl, J.-M., Heidenreich, T., & Boomgaarden, H. G. (2019). When the journey is as important as the goal: A roadmap to multilingual dictionary construction. *International Journal of Communication, 13*, 4000–4020.

Lo, S. L., Cambria, E., Chiong, R., & Cornforth, D. (2016). Multilingual sentiment analysis: From formal to informal and scarce resource languages. *Artificial Intelligence Review, 48*(4), 499–527. https://doi.org/10.1007/s10462-016-9508-4

Malmasi, S., & Zampieri, M. (2017). Detecting hate speech in social media. *RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning*, 467–472. https://doi.org/10.26615/978-954-452-049-6_062

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Manrique, N. (1999). *La piel y la pluma: Escritos sobre literatura, etnicidad y racismo [The skin and the pen: Writings on literature, ethnicity and racism]*. CiDiAG.

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930–946. https://doi.org/10.1080/1369118x.2017.1293130

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. *Proceedings of the 10th ACM Conference on Web Science*, 173–182.

McLaren, L. M. (2003). Anti-immigrant prejudice in Europe: Contact, threat perception, and preferences for the exclusion of migrants. *Social Forces*, *81*(3), 909–936. https://doi.org/10.1353/sof.2003.0038

Merkin, R. (2012). Sexual harassment indicators: The socio-cultural and cultural impact of marital status, age, education, race, and sex in Latin America. *Intercultural Communication Studies*, *21*(1).

Mittelstadt, M. (2020, August 27). *Profile of Venezuelan refugees and migrants in Latin America and the Caribbean reveals country-to-country variations in their characteristics and experiences.* IOM UN Migration. https://rosanjose.iom.int/site/en/news/profile-venezuelan-refugees-and-migrants-latin-america-and-caribbean-reveals-country-country

Müller, K., & Schwarz, C. (2017). Fanning the flames of hate: Social media and hate crime. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3082972

Murthy, D., & Sharma, S. (2019). Visualizing YouTube's comment space: Online

hostility as a networked phenomena. *New Media & Society*, *21*(1), 191–213. https://doi.org/10.1177/1461444818792393

Nuñez, A., González, P., Talavera, G. A., Sanchez-Johnsen, L., Roesch, S. C., Davis, S. M., Arguelles, W., Womack, V. Y., Ostrovsky, N. W., Ojeda, L., Penedo, F. J., & Gallo, L. C. (2016). Machismo, marianismo, and negative cognitive-emotional factors: Findings from the Hispanic community health study/study of Latinos sociocultural ancillary study. *Journal of Latina/o Psychology*, *4*(4), 202–217. https://doi.org/10.1037/lat0000050

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. In M. N. Warehime (Ed.), *Sociological studies of children and youth* (Vol. 18, pp. 253–273). Emerald Group Publishing Limited. https://doi.org/10.1108/S1537-466120140000018021

Olmos Alcaraz, A. (2018). Alteridad, migraciones y racismo en redes sociales virtuales: Un estudio de caso en Facebook [Alterity, migrations and racism in virtual social networks: A case study on Facebook]. *REMHU: Revista Interdisciplinar Da Mobilidade Humana*, *26*(53), 41–60. https://doi.org/10.1590/1980-85852503880005304

Omi, M., & Winant, H. (2015). *Racial formation in the United States* (3rd ed.). Routledge/Taylor & Francis Group.

Ooms, J. (2018). *cld3: Google's compact language detector 3* (R package version 1.1) [Computer software]. https://CRAN.R-project.org/package=cld3

Parekh, B. (2006). Hate speech. *Public Policy Research*, *12*(4), 213–223. https://doi.org/10.1111/j.1070-3535.2005.00405.x

Pelkoja, J. (2018, April 7). *Visual XGBoost tuning with caret*. Retrieved November 30, 2020, from https://kaggle.com/pelkoja/visual-xgboost-tuning-with-caret

Pew Research Center. (2014). *Religion in Latin America. Widespread change in a historically Catholic religion*. www.pewresearch.org/wp-content/uploads/sites/7/2014/11/Religion-in-Latin-America-11-12-PM-full-PDF.pdf

Plummer, D. C. (2001). The quest for modern manhood: Masculine stereotypes, peer culture and the social significance of homophobia. *Journal of Adolescence*, *24*(1), 15–23. https://doi.org/10.1006/jado.2000.0370

Radkiewicz, P. (2003). The national values as a concept helpful in explaining the development of nationalistic attitudes and xenophobia. *Polish Psychological Bulletin*, *34*(1), 5–14.

Ramon-Berjano, C. (2011). Regional integration in Latin America: MERCOSUR, UNASUR and ZICOSUR. *Regions Magazine*, *281*(1), 10–10. https://doi.org/10.1080/13673882.2011.9672722

Ramos, J. (2003). *Using TF-IDF to determine word relevance in document queries*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf.

Ramsey, G., & Sánchez-Garzoli, G. (2018). *Venezuela's migration and refugee crisis as seen from the Colombian and Brazilian borders*. Washington Office on Latin America. https://reliefweb.int/sites/reliefweb.int/files/resources/Final-VZ-Migration-Report-Final_1.pdf

R Core Team. (2018). *R: A language and environment for statistical computing*. The R Project for Statistical Computing. https://www.R-project.org/

Rinker, T. W. (2017). *qdapRegex: Regular expression removal, extraction, and replacement tools* (version 0.7.2) [Computer software]. http://github.com/trinker/qdapRegex

Rinker, T. W. (2018). lexicon: Lexicon data (Version 1.1.3) [Computer software]. http://github.com/trinker/lexicon

Ripoll, S., & Navas-Alemán, L. (2018). *Xenofobia y discriminación hacia refugiados y migrantes venezolanos en Ecuador y lecciones aprendidas para la promoción de la inclusión social [Xenophobia and discrimination against Venezuelan refugees and migrants in Ecuador and lessons learned for the promotion of social inclusion]*. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14130

Rivero, P. (2019). Sí, pero no aquí: Percepciones de xenofobia y discriminación hacia migrantes de Venezuela en Colombia, Ecuador y Perú *[Yes, but not here: Perceptions of xenophobia and discrimination against Venezuelan migrants in Colombia, Ecuador and Peru]*. Oxfam International. https://doi.org/10.21201/2019.5303

Rocha, N. (2018, October 10). *Brotes de xenofobia en Latinoamérica, un problema de identidad [Outbreaks of xenophobia in Latin America, a problem of identity]*. https://unperiodico.unal.edu.co/pages/detail/brotes-de-xenofobia-en-latinoamerica-un-problema-de-identidad/

Rodríguez García, H. (2011). Mestizaje y conflictos sociales. El caso de la construcción nacional boliviana [Miscegenation and social conflicts. The case of Bolivian national construction]. *Del Mestizaje a la Hibridez: Categorías Culturales en América Latina*, *8*(9), 145–182. https://revistas.ucr.ac.cr/index.php/intercambio/article/view/2204

Saa, I. L., Novak, M., Morales, A. J., & Pentland, A. (2020). Looking for a better future: Modeling migrant mobility. *Applied Network Science*, *5*(1), 70. https://doi.org/10.1007/s41109-020-00308-9

Sahhar, G. (2021). *¿Cuáles son los países que exigen visa a los venezolanos? [Which countries require a visa for Venezuelans?]*. https://eldiario.com/2021/01/17/paises-exigen-visa-venezolanos/

Sayimer, İ., & Derman, M. R. (2017). Syrian refugees as victims of fear and danger discourse in social media: A YouTube analysis. *Global Media Journal TR Edition*, *8*(15) , 384–403. https://globalmediajournaltr.yeditepe.edu.tr/sites/default/files/19_idil_sayimer_malgorzata_rabenda_derman.pdf

Schindler, M., & Domahidi, E. (2021). The growing field of interdisciplinary research on user comments: A computational scoping review. *New Media & Society, 1*–19. https://doi.org/10.1177/1461444821994491

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. https://doi.org/10.18653/v1/w17-1101

Schultes, P., Dorner, V., & Lehner, F. (2013). Leave a comment! An in-depth analysis of user comments on YouTube. *Wirtschaftsinformatik Proceedings*, *42*, 659–673. https://aisel.aisnet.org/wi2013/42.

Spertus, E. (1997). Smokey: Automatic recognition of hostile messages. *IAAI-97 Proceedings,* 1058–1065. *American Association of Artificial Intelligence*. https://www.aaai.org/Papers/IAAI/1997/IAAI97-209.pdf

Top sites. (2019). Alexa. Retrieved November 30, 2020, from www.alexa.com/topsites

Torres-Moreno, J. M. (2012). Beyond stemming and lemmatization: Ultra-stemming to improve automatic text summarization. *ArXiv*, *abs/1209.3126*, 22. https://arxiv.org/abs/1209.3126v1

UN Educational, Scientific and Cultural Organization. (2017). *Xenophobia.* International migration. Retrieved November 30, 2020, from www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/xenophobia/

UN High Commissioner for Refugees. (2010). *Convention and protocol relating to the status of refugees*. UNHCR. www.unhcr.org/protection/basic/3b66c2aa10/convention-protocol-relating-status-refugees.html

UN High Commissioner for Refugees. (2019). *Refugees and migrants from Venezuela top 4 million: UNHCR and IOM*. UNHCR. www.unhcr.org/news/press/2019/6/5cfa2a4a4/refugees-migrants-venezuela-top-4-million-unhcr-iom.html

Unicode. (2019). *Emoji keyboard/display test data for UTS #51* [Data set]. Retrieved November 30, 2020, from https://unicode.org/Public/emoji/12.0/emoji-test.txt

Economic Commission for Latin America and the Caribbean. (2013). *Three decades of uneven and unstable growth* (65th ed.). United Nations. http://hdl.handle.net/11362/1086.

Welsh, T. (2018, September 19). *Venezuela crisis is "on the scale of Syria," UNHCR says*. Devex. https://www.devex.com/news/venezuela-crisis-is-on-the-scale-of-syria-unhcr-says-93465

Wickham, H. (2019). *stringr: Simple, consistent wrappers for common string operations* (R package version 1.4.0) [Computer software]. https://CRAN.R-project.org/package=stringr

World Health Organization & World Bank. (2011). *World report on disability*. World Health Organization. https://www.who.int/publications/i/item/9789241564182

Zick, A., Wolf, C., Küpper, B., Davidov, E., Schmidt, P., & Heitmeyer, W. (2008). The syndrome of group-focused enmity: The interrelation of prejudices tested with multiple cross-sectional and panel data. *Journal of Social Issues*, *64*(2), 363–383. https://doi.org/10.1111/j.1540-4560.2008.00566.x