The Geopolitical Repercussions of US Anti-immigrant Rhetoric on Mexican Online Speech About Migration: A Transdisciplinary Approach

Thomas Cattin, Alejandro Molina-Villegas, Julieta Fuentes-Carrera, and Oscar S. Siordia

Abstract This paper presents an ongoing research project that aims to propose a geopolitical analysis of anti-immigrant speech published on the Mexican twitosphere. While Mexico has long defined itself as an emigration country, the apparent growing presence of anti-immigrant discourse online, especially at the Mexican borders, invites us to question the impact of Americans' anti-immigrant speech, bolstered by Donald Trump's election and presidency, on Mexicans' representations. We thus propose a transdisciplinary approach that combines a Convolutional Neuronal Network to detect anti-immigrant speech in geolocalized tweets in Mexican Spanish and a geopolitical diachronic analysis to estimate the relationship between such speeches and Americans' anti-immigrant online representations. With an overall accuracy of 0.76, we are confident that with some improvements the CNN model will be able to detect Mexican anti-immigrant speech on Twitter. We finally discuss that the scope of the analysis would be greatly improved if paired with network and territorial analysis of Mexican anti-immigrant tweets.

Keywords Geopolitical repercussions · United States · Mexico · Twitter · Georeferenced anti-immigrant speech · Hate speech detection

T. Cattin (✉)
Centro de Investigación en Ciencias de Información Geoespacial, Mexico and Institut Français de Géopolitique, Mecico City, France
e-mail: tcattin@centrogeo.edu.mx

A. Molina-Villegas
Consejo Nacional de Ciencia y Tecnología, Centro de Investigación en Ciencias de Información Geoespacial, Mecico City, Mexico
e-mail: amolina@conacyt.mx

J. Fuentes-Carrera
Centro de Investigación en Ciencias de Información Geoespacial, Mecico City, Mexico

O. S. Siordia
Laboratorio Nacional de GeoInteligencia, Centro de Investigación en Ciencias de Información Geoespacial, Mecico City, Mexico

R. Tapia-McClung et al. (eds.), *Advances in Geospatial Data Science*, Lecture Notes in Geoinformation and Cartography, https://doi.org/10.1007/978-3-030-98096-2_4

1 Introduction

The arrival of several thousands of asylum seekers from Central America at the border town of Tijuana, in November 2018, showed anti-immigration attitudes in a part of the local population. The Mexican anti-immigrant speech was evident online, such as in a Facebook group *Tijuana en contra de la caravana migrante* with up to 4000 members (UnionTJ664 2018), and was instrumentalized by part of the local political elite. The former mayor of Tijuana, Juan Manuel Gastélum Buenrostro, was seen wearing a *Make Tijuana Great Again* cap.

Mexican anti-immigrant speech becoming visible is a problem to be addressed in geopolitical terms: The Mexican territory is the main migratory terrestrial interface for clandestine migrations towards the United States (Casillas 2008) (Fig. 1). Since the September 11th, 2001, terrorist attacks, the Mexican borders have become geostrategic externalization territories for the United States border controls. This was done by joint policies alongside the Mexican government, such as the Southern Border Program (*Programa Frontera sur*) in 2014. All while the clandestine migration notably from Central America increased (Capps et al. 2019), these policies have transformed the Mexican borders into migratory buffer zones where large groups of migrants get trapped for several months waiting for the resolution of their migratory processes.

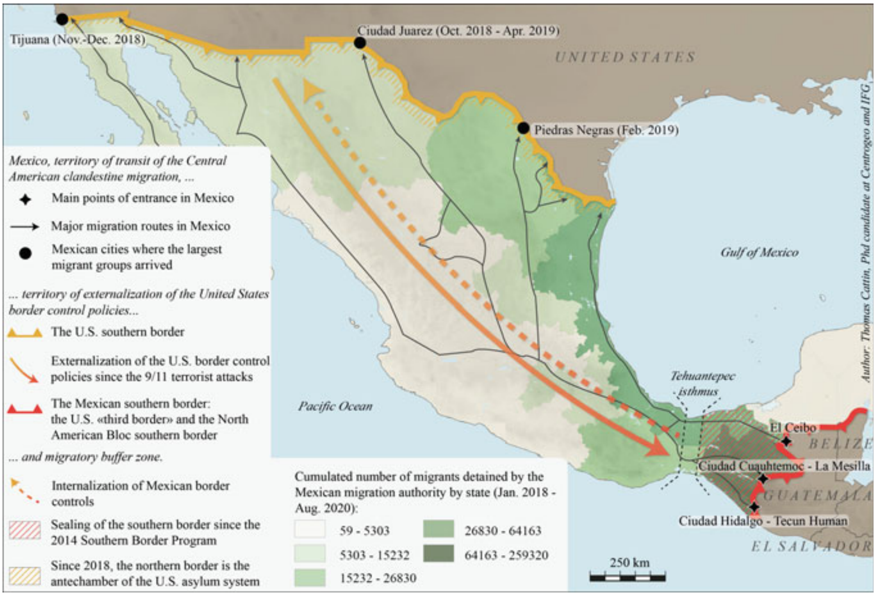

Fig. 1 Mexico: migration interface and migratory buffer zone of the United States (Casillas 2008; UPM 2020; Zepeda and Fuentes Carrera 2020)

Mexico does not yet seem to have the actors and mechanisms favoring the emergence of intense and lasting anti-immigrant speech, such as the political instrumentalization of the subject by opposed partisan blocs (Ernst et al. September 2017). However, expressions such as *"Make Tijuana Great Again"* or the hashtag *#MexicoPrimero*, derived from Donald Trump's campaign slogans which aimed to electorally exploit inter-ethnic power rivalry in the US, indicate some interiorization of American anti-immigrant rhetoric in Mexico. Such interiorization invites us to think about Mexican anti-immigrant speech not as a cultural but as a geopolitical phenomenon where the migratory situation in Mexico, a byproduct of US migration policies, and the electoral power struggle in the US have bolstered the online transnationalization of American anti-immigrant speech.

There are two objectives to this paper. First, to detail the approach used to build a Convolutional Neuronal Network (CNN) model able to detect anti-immigrant speech in geolocalized Mexican Spanish tweets and to present the preliminary result of said model. Second, to propose a transdisciplinary methodology, based on the data obtained from the classification, to verify the hypothesis that American anti-immigrant speech online during Donald Trump's presidency has had a quantitative and qualitative impact on Mexican anti-immigrant tweets.

2 State of the Art

2.1 State of the Art of the Online Analysis of Anti-immigrant Speech

With the emergence of the World Wide Web, social media, and in particular, Twitter, have imposed themselves as privileged media platforms for verbal opposition, especially on migration issues, between users that interact almost instantaneously through the content they produce (Karatzogianni et al. 2016; Koylu et al. 2019). An essential takeaway from these studies is the transnational aspect of the debate on social media (Ferra and Nguyen 2017; Nguyen 2016). Toudert recently analyzed the conflictual nature of the debate on Twitter about the arrival in Tijuana of the Central American migrant caravan (Toudert 2021). In most works, the sentiment analysis is limited to surface forms, such as lexicon and syntax. We argue that developing a classification model based on deep learning techniques can improve the detection of anti-immigrant speech. Furthermore, the use of geolocalized data from Twitter will enable us to isolate and analyze specifically the Mexican anti-immigrant speech.

2.2 State of the Art of Online Hate Speech Detection

The recurrent presence of hate speech on social media has generated an interest in the automatic detection of this type of content by using Natural Language Processing (NLP) techniques (Fortuna and Nunes September 2018). Recent works have demonstrated the efficiency of an approach based on neuronal networks to detect hate speech in English (Ridenhour et al. 2020; Zhang et al. 2018) and Spanish (Pereira-Kohatsu et al. 2019). Other research has focused on certain types of hate speech such as misogyny (Molina-Villegas 2021). Even if the detection of anti-immigrant hate speech in Spanish is of undeniable interest (Basile et al. 2019), this domain is still unexplored. This project aims at building a model able to detect anti-immigrant speech in Spanish in the Mexican territorial context.

3 Anti-immigrant Speech Detection Using CNN

3.1 Twitter Georeferenced Data and How to Obtain It

Twitter data presents the considerable advantage of being relatively easy to access with the platforms' Application Programming Interfaces. The basic data unity is the tweet, a short message (280 characters) accompanied by metadata such as username, date, language, geolocalization, and information about the interaction of other users (retweets, comments, mentions, and favorites). In our case, the detection of anti-immigrant hate speech comes from Machine Learning models specifically created to analyze patterns in the textual information of the previously collected tweets based on geolocalization criteria.

A big data set of tweets published in Mexico between January 2017 and May 2021 were collected using the *Autómata Geointeligente en Internet* (AGEI) developed by Centrogeo[1] and previously used for another research work (López-Ramírez et al. 2019). Due to the fact that only about 1 to 2 percent of all tweets are georeferenced, and about 1 percent of all tweets contain hate speech (Pereira-Kohatsu et al. 2019), the percentage for anti-immigrant speech is probably much lower, it is necessary to dispose of a large raw database.

Once downloaded, the AGEI data was mapped in a Geographical Information System (GIS) and regrouped by year, three-month periods, and by region of publication (Fig. 2). These regions correspond to northern Mexican border states, the southern border states, the rest of the Mexican states, and the US southern border region. To remediate possible imprecisions in tweets' geolocalisation, we applied a buffer of 20km to the four regions.

[1] https://www.centrogeo.org.mx/geointeligencia.

3.2 Manual Tweet Labeling

The first step to build a supervised classification model is the most time-consuming one: labeling a large number of tweets manually into our two classes that are anti-immigrant and those that are not (Fig. 2). However, to speed up this process, the tweets collected with AGEI were filtered using a catalog of 49 words and expressions susceptible to anti-immigrant speech. In parallel, a Twitter search by context and users was programmed, a tweet published in a specific context or by a specific user where interactions with the tweet could contain anti-immigrant speech.

The tweets were then manually labeled according to five criteria. These criteria were elaborated from the geopolitical conceptualization of anti-immigrant speech. The criteria were sufficiently simple to allow non-specialists to use them and general enough to detect anti-immigrant speech published on Twitter in Mexico. The criteria were qualitatively validated by a panel of experts. To further improve the objectivity of criteria, it is planned to measure statistically the inter-rater agreement based upon experts' individual classification, using the criteria, of a small data set containing both anti-immigrant and not anti-immigrant tweets.

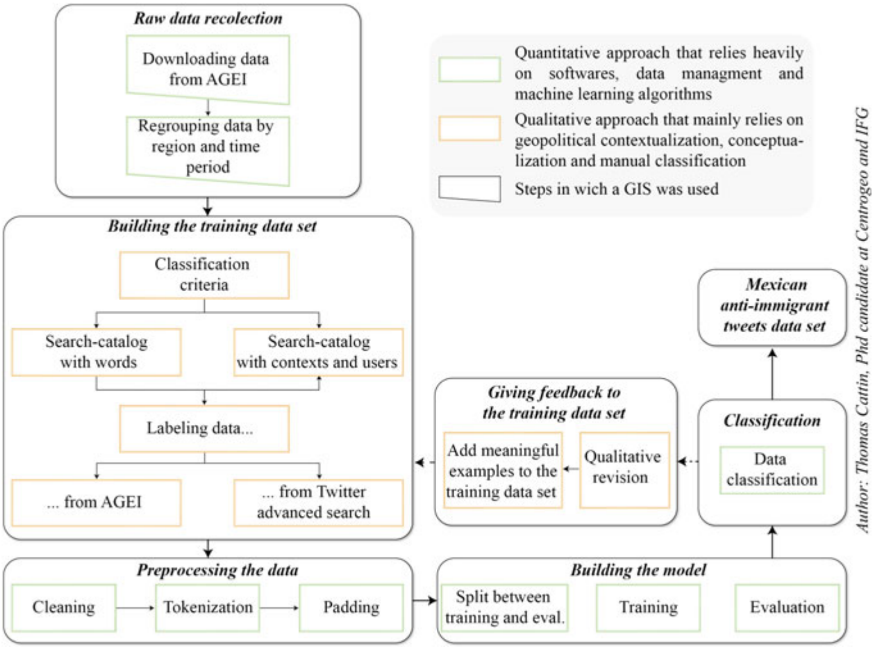

Fig. 2 Methodology for building the model

3.3 Model Training and Classification

Before training the model and classifying the data, they were submitted to a series of transformations to generate representative features, which served as inputs in the model (Fig. 2). These operations were done in the following order: data cleaning, tokenization, and padding. Labeled and pretreated data were then divided into two data sets to train and evaluate the model. Because of the imbalance between negative (4548) and positive (1073) tweets in our training data, 18.5 percent of the positive tweets (200), and the same number of negative tweets, were used for the evaluation in order to measure how accurate the model is for both classes.

The training data set was then used to train a Convolutional Neuronal Network (Fig. 3). This is a technique primarily used in image processing, but applied to texts, a CNN is very effective in applying successive filters that reduce the entry vectors' dimensions (embeddings) while increasing the effectiveness of finding text patterns (parameters). This type of model has proven efficient in hate speech detection (Molina-Villegas 2021; Zhang et al. 2018) and Geographic Named Entity Recognition (Molina-Villegas et al. 2021).

Once trained, the model was used to classify pretreated data collected from AGEI. The earliest classification was used as feedback for the training stage (Fig. 2). The qualitative revision of the classified data proved useful for adding negatives and positives to the training data set by manually reclassifying false positives and false negatives.

3.4 Results on Classification Model not Anti-immigrant, Anti-immigrant

The Table 1 presents the preliminary results for the statistical performance of the binary classification model (not anti-immigrant, anti-immigrant). The average precision for the not anti-immigrant class indicates that we still have a relatively high number of false positives, which was corroborated during the qualitative revision of the classification output. Whereas the recall for the anti-immigrant class indicates that the model fails to detect a good portion of the anti-immigrant tweets.

To further improve the model, we need to increase the training data set with both anti-immigrant and not anti-immigrant tweets. Our current training data set contains

Table 1 Results of binary classification for anti-immigrant speech in Mexican tweets

	Accuracy	Precision	Recall	F-measure	support
Not anti-immigrant		0.7219	0.8700	0.7891	200
Anti-immigrant		0.8364	0.6650	0.7409	200
All	0.7675				400

only 1073 anti-immigrant tweets and mainly consists of data from the northern and southern border regions in 2018 and 2019. It is therefore neither geographically nor temporally representative of the entire data set. This could explain the high number of false positives found during a qualitative revision of newly classified data. It will also be necessary to experiment with the variation of specific parameters of the model, such as the number of filters (300), the number of neurons in the hidden layer (300), and the dropout rate (0.45).

4 A Geopolitical Analysis of Mexican Anti-immigrant Speech on Twitter During the Trump Administration

It now has to be said that information obtained from Twitter does not represent the whole anti-immigrant speech in Mexico. Those who publish on Twitter have internet access for which Mexico has a marked urban/rural fracture (INEGI 2020), and they are relatively politicized as Twitter appears to be the preferred platform for interactions between the ruling class and the governed (Bruns and Stieglitz 2014; Tandoc andJohnson 2016). Furthermore, georeferenced tweets represent a tiny fraction of all tweets, and those who publish geographical data are not representative of the wider Twitter population (Sloan and Morgan 2015). In this section, we do not pretend to extend the analyses outside of Twitter for those reasons.

In recent years, Donald Trump's ability to steer the terms of the migration debate on Twitter has been demonstrated (Koylu et al. 2019). Through his @realDonaldTrump Twitter account, he instrumentalized the core elements of American racist discourse now aimed at those who come from beyond the southern US border (Haney-López 2014). We thus propose to use @realDonaldTrump's anti-immigrant tweets to estimate by proxy the influence of US anti-immigrant discourse on Mexican anti-immigrant speech (Fig. 3).

In geopolitics, the diachronic analysis revolves around the division of time into crises and ruptures that highlight the evolution of a phenomenon on the territory. We can transpose this analysis method on Twitter by dividing the publication history of @realDonaldTrump into temporal milestones, tweets that crystallize the evolution of Donald Trump's discourse on migration. The @realDonaldTrump tweets published between January 2017 and January 2021 were downloaded from the Trump Twitter Archive[2] and the ones related to migration were manually selected. Based on personal knowledge, we then determine the temporal milestones for Trump's domestic and international migration tweets.

Once validated, those temporal milestones will be used as an analysis grid for Mexican anti-immigrant tweets obtained from the classification (Fig. 3). We established two ways to infer repercussions of American anti-immigrant speech on those tweets.

[2] https://www.thetrumparchive.com.

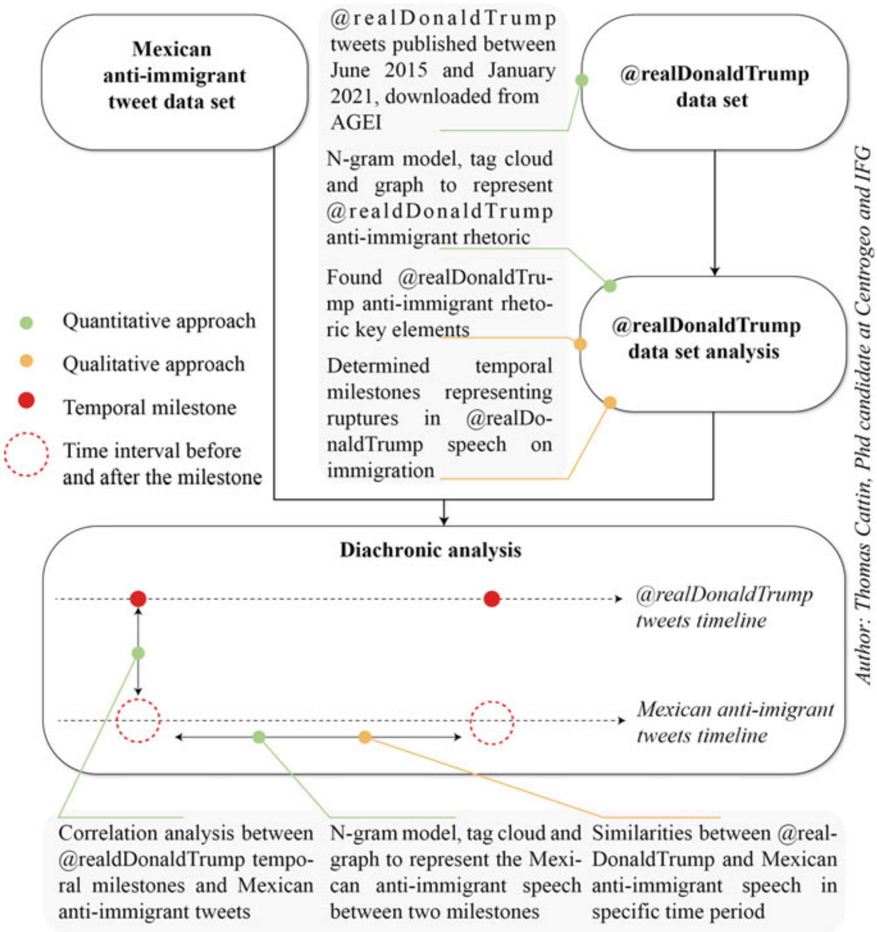

Fig. 3 Roadmap for diachronic analysis

Firstly, by examining over the period the correlation between the @realDonaldTrump milestones and the variations in the number of Mexican anti-immigrant tweets. Secondly, by looking for lexical similarities between Mexican speech and @realDonaldTrump's using NLP techniques like n-gram models. With those two analyses, we should have a somewhat accurate picture of how American anti-immigrant rhetoric, Donald Trump's or similar rhetoric used by users unknown to us for the moment, impacted both quantitatively and qualitatively the Mexican anti-immigrant speech on Twitter.

5 Conclusions

When completed, this research project will make several contributions to both NLP and geopolitics communities. In addition to providing a concrete application of NLP techniques in social sciences, we will have designed a data base and a model for detecting anti-immigrant speech in Spanish. Regarding geopolitics, this work will provide evidence that, through digital social networks, highly politicized and sensitive debates in one territory can be reshaped by the repercussions of similar debates in entirely different territorial contexts. Finally, we hope that this project will serve as an example of how useful a transdisciplinary approach can be when the exponential growth of available data and the emergence of the internet as a real digital space offer an unprecedented opportunity to deepen our knowledge about social and territorial interactions.

There is still a lot to be done and we argue that this research would be greatly improved if combined with other analyses. Network analyses based on graph theory techniques have proved successful in studying interactions between digital social network users and their consequences (Gonzólez-Bailón 2017), including in geopolitics (Gérard and Marotte 2020). Based on interactions such as mentions, replies, or favorites, we could reconstruct the network that underlies the Mexican anti-immigrant speech on Twitter. With a topographic analysis of this network, we could identify the main users and communities of users who disseminate those speeches and how they bridge the boundaries of the Mexican and American national debate about migration.

One of the major contemporary geopolitical questions is to understand how interactions within digital spaces impact the production of territories and *vice versa*. To investigate this relationship, we could use the network analysis as an input for two field studies, one in Tijuana (northern Mexico border) and the other in Tapachula (southern Mexico border). Where possible, local actors of those two border cities whose digital *alter ego* plays a role in the Mexican anti-immigrant community on Twitter, and whom we manage to identify, would be questioned about their views on immigration and their presence online during qualitative interviews.

References

Basile V, Bosco C, Fersini E, Debora N, Patti V, Pardo FM, Rosso P, Sanguinetti M (2019) SemEval-2019 task 5: multilingual detection of hate speech against immigrants and women in twitter. In: Proceedings of the 13th international workshop on semantic evaluation, pp 54–63, Minneapolis, Minnesota, USA, 2019. Association for Computational Linguistics

Bruns A, Stieglitz S (2014) Twitter data: what do they represent? Inf Technol 56(5):240–245

Capps R, Meissner D, Ruiz S, Ariel G, Bolter J, Pierce S (2019) From control to crisis: changing trendrs and policies reshaping U.S.-Mexcio border enforcment. Technical report, Migration Policy Institute, Washington, DC, 2019

Casillas R (2008) Las rutas de los centroamericanos por México, un ejercicio de caracterización, actores principales y complejidades. Migración y desarrollo 10:157–174

Ernst N, Engesser S, Büchel F, Blassnig S, Esser F (2017) Extreme parties and populism: an analysis of Facebook and Twitter across six countries. Inf Commun Soc 20(9):1347–1364

Ferra I, Nguyen D (2017) #migrantcrisis: & "tagging" the European migration crisis on twitter. J Commun Manag 21(4):411–426

Fortuna P, Nunes S (2018) A survey on automatic detection of hate speech in text. ACM Comput Surv 51(4):1–30

Gonzólez-Bailón S (2017) Decoding the social world: data science and the unintended consequences of communication. Information policy series. MIT Press, Cambridge, MA

Gérard C, Marotte G (2020) #AffaireBenalla : déconstruction d'une polémique sur le rôle de la communauté Twitter ≪ russophile ≫ dans le débat politique français. Hérodote, N ∘ 177–178(2):125

Haney-López I (2014) Dog whistle politics: how coded racial appeals have reinvented racism and wrecked the middle class. Oxford University Press, Oxford. OCLC: 884873319

INEGI (2020) En méxico hay 84.1 millones de usuarios de internet y 88.2 millones de usuarios de teléfonos celulares: Endutih, (2020) Technical report. Mexico, Instituto Nacional de Estadística y Geografía, Mexico City, p 2021

Karatzogianni A, Nguyen D, Serafinelli E (eds) (2016) The digital transformation of the public sphere: conflict, migration, crisis and culture in digital networks. Palgrave Macmillan UK

Koylu C, Larson R, Dietrich BJ, Lee K-P (2019) Carsentogram: geovisual text analytics for exploring spatiotemporal variation in public discourse on twitter. Cartogr Geogr Inf Sci 46(1):57–71

López-Ramírez P, Molina-Villegas A, Siordia OS (2019) Geographical aggregation of microblog posts for LDA topic modeling. J Intell Fuzzy Syst 36(5):4901–4908

Molina-Villegas A (2021) La incidencia de las voces misóginas sobre el espacio digital en méxico. In: Pérez-Barajas AE, Arellano-Ceballos AC (eds) Jóvenes, plataformas digitales y lenguajes: diversidad lingüística, discursos e identidades. Elementum, Mexico (in press)

Molina-Villegas A, Muñz-Sanchez V, Arreola-Trapala J, Alcóntara F (2021) Geographic named entity recognition and disambiguation in mexican news using word embeddings. Expert Syst Appl 176:114855

Nguyen D (2016) Analysing transnational web spheres: the European example during the Eurozone Crisis. Palgrave Macmillan UK, London, pp 211–233

Pereira-Kohatsu JC, Quijano-Sónchez L, Liberatore F, Camacho-Collados M (2019) Detecting and monitoring hate speech in twitter. Sensors 19(21):4654

Ridenhour M, Bagavathi A, Raisi E, Krishnan S (2020) Detecting online hate speech: approaches using weak supervision and network embedding models. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M (eds) Social, cultural, and behavioral modeling. Series title: lecture notes in computer science, vol 12268. Springer International Publishing, Cham, pp 202–212

Sloan L, Morgan J (2015) Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. PLoS ONE 10(11)

Tandoc EC, Johnson E (2016) Most students get breaking news first from Twitter. Newspaper Res J 37(2):153–166; Publisher: SAGE Publications Inc

Toudert D (2021) Crisis de la caravana de migrantes: algunas realidades sobre el discurso público en Twitter. Migr Int 12

UnionTJ664 (2018) Tijuana en contra de la caravana migrante

UPM (2020) Eventos de extranjeros presentados ante la autoridad migratoria, según entidad federativa y municipio (2020) Technical report. Mexico, UPM-Secretaría de Gobernación, Mexico City, p 2021

Zepeda B, Fuentes-Carrera J (2020) La frontera México-Guatemala y el perímetro de seguridad de Estados-Unidos 2000-2020. In: Fuentes Carrera J (ed) Entre lo político y lo espacial: representaciones geopolíticas de la región transfronteriza México-Guatemala. Región Transfronteriza México Guatemala, pp 49–84

Zhang Z, Robinson D, Tepper J (2018) Detecting hate speech on twitter using a convolution-GRU based deep neural network. In: Gangemi A, Navigli R, Vidal M-E, Hitzler P, Troncy R, Hollink L, Tordai A, Alam M (eds) The semantic web. Springer International Publishing, Cham, pp 745–760