



# ALAP 2020

IX Congreso de la Asociación  
Latinoamericana de Población



9 a 11 diciembre

EL ROL DE LOS ESTUDIOS DE POBLACIÓN TRAS LA PANDEMIA DE COVID-19 Y  
EL DESAFÍO DE LA IGUALDAD EN AMÉRICA LATINA Y EL CARIBE

*Suárez-Gutiérrez, Manuel, Universidad Veracruzana, mansuarez@uv.mx*

*Sánchez-Cervantes, José Luis, Instituto Tecnológico de Orizaba, jlsanchez@conacyt.mx*

*Paredes-Valverde, Mario Andrés, Instituto Tecnológico de Teziutlán, mapv1015@gmail.com*

*Marín Lozano, Erasto Alfonso, Universidad Veracruzana, emarin@uv.mx*

*Guzmán Coutiño, Héctor, Universidad Veracruzana, hguzman@uv.mx*

Percepción social sobre violencia en México desde la red  
social online de Twitter

## Resumen

Medir de forma cualitativa la percepción social sobre la violencia en México a partir de la red social online de Twitter conlleva a entrañar el lado oculto de esta plataforma a través de los metadatos. Se expone la relevancia de identificar técnicas novedosas para un tema que aqueja a la sociedad mexicana desde hace algún tiempo y que, además, aporten a comprender desde otra perspectiva el sentir social. La metodología implementada retoma tecnologías emergentes basadas en Big Data, semántica web, minería de datos web y análisis de redes sociales. Los resultados ilustran a través de un análisis de hashtags y por grafos la percepción social derivada de los tweets vinculados a temas de violencia.

**Palabras Clave:** Percepción Social; Violencia; Twitter

## Abstract

Qualitatively measure the social perception of violence in Mexico from the online social network of Twitter, leads to scrutinize the hidden side of this site through the metadata. We expose the relevance of identifying novelty techniques for an issue that worries for some time to Mexicans, and also, they contribute to understanding the social perception from another angle. The methodology implemented, is based on emerging technologies related to Big Data, Semantic Web, Web mining, and Social Media. Results show the social perception through a hashtag analysis and graph analysis, from tweets linked to violence topics.

**Keywords:** Social perception; Violence; Twitter

## Introducción

El presente estudio es un producto que forma parte del proyecto de investigación denominado “Análisis de información proveniente de la red social online de Twitter, utilizando técnicas de Big Data”. De este proyecto se deriva la tesis doctoral “Grado de percepción social sobre la violencia en México a partir de datos difundidos en la red social online de Twitter”, la cual tiene por objeto generar un índice que logre cuantificar a partir de metadatos de tipo cualitativos, la percepción social de la violencia en México a una escala de entidades federativas y zonas metropolitanas.

Hablar de percepción social conlleva a indicar la influencia de factores sociales y culturales sobre percepción y cognición. La percepción identifica a las personas visualizando el conocimiento de los otros, formación de impresiones y procesos de distribución (Santoro, 2014). Se explora la influencia de los entornos, así como de situaciones sociales y culturales específicas, para emitir un juicio acorde con la investigación realizada. Del mismo modo, la percepción social también incluye elementos para el reconocimiento de emociones, el sentimiento de los individuos en su medio físico y social (Santoro, 2014).

Considerar la percepción de las personas a partir de datos provenientes de la red social online de Twitter, permite hacer inferencias sobre las características (observables y no observables) de los usuarios que publican en dicha red social. Las características observables son los metadatos de cada usuario, los tweets publicados y reenviados. Por otro lado, las características no observables corresponden al sentimiento semántico del tweet, la manifestación de conductas e ideologías; en general son aquellos elementos que dilucidan la manera de pensar de un usuario a partir de sus publicaciones.

La percepción social crea expectativas de credibilidad respecto a las conductas sociales manifestadas en una red social online y supone la existencia de una interpretación de la realidad social a partir de

los tweets (esto es pasar de lo virtual a la realidad) (Vaidya et al., 2019). Se identifica la influencia de factores sociales y culturales, así como la forma en que el medio social afecta los procesos perceptuales. Estos procesos perceptuales deben evitar caer en prejuicios a partir de la generación de estereotipos, ya que se puede caer en un sesgo o prejuicio, al catalogar a un usuario por alguna circunstancia externa y dejar de lado sus cualidades observables (Santoro, 2014).

Derivado de la creciente incidencia delictiva en México, se decidió recabar información proveniente de la red social online de Twitter, para analizar los metadatos de las publicaciones realizadas a fin de obtener la percepción de los twittereros sobre el tema de violencia. En otras palabras, se busca caracterizar un problema que aqueja no solo a la sociedad mexicana, sino también a varias latitudes de América Latina, desde una perspectiva diferente, esto es, desde la percepción de los twittereros.

Para poder hacer este análisis de los metadatos, se requiere el uso de técnicas basadas en Big Data, Twitter Analytics, Web Mining y Semantic Web. Por ello, se muestra el diseño de un modelo para la adquisición de datos desde la red social online de Twitter.

Si bien el estudio abarca únicamente la geopolítica de México, la metodología propuesta permite que en futuras investigaciones con apoyo de colegas de otros países se pueda ampliar la muestra a diferentes latitudes geoespaciales.

Dentro de los resultados obtenidos se muestra: la correlación de los hashtags (#) ordenados por clústeres de palabras clave; y un análisis por grafos, identificando la importancia de los temas emanados desde la agenda de la política del gobierno federal mexicano con respecto al tema de vulnerabilidad social y violencia en México.

## **Planteamiento del problema**

Nos encontramos en la cúspide de un cambio de paradigma en el estudio de la comunicación y del intercambio de ideas de la sociedad. En donde, con el surgimiento de las redes sociales online, se dotó de una libertad de expresión a la sociedad, en donde cualquiera con acceso a internet puede difundir sus ideas en estas plataformas digitales. Libertad de expresión que llega a comprometer algunos gobiernos, ya que a través de movimientos sociales digitales pueden generar tendencias que incomoden a la agenda pública. Tal es el caso, que se ha llegado a dialogar sobre la restricción y control de lo que se publica, e incluso en censurar algunos temas.

Por otro lado, las redes sociales online han servido para catapultar temas de la agenda pública del mismo gobierno, a partir de posicionar temas de tendencias bombardeando de información a los usuarios de las redes sociales online. Con ello, se coloca en el diálogo popular los temas que el mismo gobierno desea que se comenten en estos medios de difusión masivo digitales.

Comprender el contexto mexicano que ha llevado a la violencia e inseguridad a ser un tema en el diálogo cotidiano en la sociedad es complejo; mencionar cuando comenzó este problema a socavar a la sociedad y a ser objeto de preocupación lo es aún más complicado. De acuerdo con los datos estadísticos emitidos por el INEGI(2019) a través de la Encuesta Nacional de Victimización y Percepción Sobre Seguridad Pública (ENVIPE), desde el 2011 se convirtió en información de interés nacional, por lo que, a partir de esta fecha, se cuenta con datos a una escala de entidades federativas de forma regular y periódica.

Ante este panorama, se detectó que la red social online de Twitter ha tomado un interés particular en temas relacionados con la difusión de acontecimientos bélicos, políticos, electorales, educativos, entre otros (Bolgov et al., 2017; Ristea et al., 2017; Sandoval-Almazan et al., 2014; Sandoval-Almazan & Gil-Garcia, 2013; Suárez-Serrato et al., 2018). Por lo que, se convierte en una herramienta

relevante para comprender lo que acaece la sociedad desde una mirada a partir de un espacio de comunicación virtual.

## Objetivo

Determinar la percepción social a una escala de clústeres de palabras clave sobre el tema de la violencia en México, a partir de los tweets difundidos en la red social online de Twitter con el propósito de comprender un aspecto novedoso de la interacción social desde un medio de comunicación virtual.

## Metodología

Se cuenta con información recabada entre el 14 de mayo y el 27 de junio de 2019 desde la red social online de Twitter. El tiempo de captura se decidió a partir de dejar pasar los primeros 100 días del gobierno federal entrante, a manera de evitar que se capturaran sentimientos como la euforia y el malestar de los twitters por la transición gubernamental. Sentimientos que infieren directamente en el análisis de los datos a obtener.

### *Procedimiento Experimental*

- **Selección de palabras clave:** Se identifican a usuarios destacados con respecto a la temática de violencia en México. Un usuario en Twitter puede ser una empresa, una organización sin fines de lucro, una institución pública o bien un ciudadano. Para el caso de este estudio, se utilizaron cuentas de usuarios provenientes de las instituciones gubernamentales, ya que son perfiles que pueden ser verificados. En este caso, se seleccionaron 60 cuentas de usuarios de instituciones gubernamentales a nivel de estatal y municipal; y 4 cuentas a nivel nacional (ver Tabla 1).

Con la herramienta de TweetReach (Union Metrics, 2020) se analizó a cada uno de los usuarios seleccionados para identificar las palabras clave y hashtags con mayor frecuencia de cada uno de ellos. Resultando un listado de 150 términos clave (ver Tabla 2) utilizados para la extracción de datos generados en la red social online de Twitter.

Tabla 1. Cuentas de usuario de la red social online de Twitter seleccionadas.

Nombre	Twitter	Ámbito
C4 Baja California	<a href="#">@911_C4BC</a>	Estatal
Emergencias 911 Jalisco	<a href="#">@911Jalisco</a>	Estatal
Emergencias 911 Puebla	<a href="#">@911Puebla</a>	Estatal
AIC PGJ Guanajuato	<a href="#">@AIC_Guanajuato</a>	Estatal
AMIC Sonora	<a href="#">@Amicsonora</a>	Estatal
C5 CDMX	<a href="#">@C5_CDMX</a>	Estatal
C5 Estado de México	<a href="#">@C5Edomex</a>	Estatal
Fiscalía Durango	<a href="#">@DgoFiscalia</a>	Estatal
FGE Coahuila	<a href="#">@FGECoahuila</a>	Estatal
Fiscalía Estado de México	<a href="#">@FiscaliaEdomex</a>	Estatal
Fiscalía de Jalisco	<a href="#">@FiscaliaJal</a>	Estatal
Fiscalía de Nayarit	<a href="#">@FiscaliaNayarit</a>	Estatal
FGE Puebla	<a href="#">@FiscaliaPuebla</a>	Estatal
PESP Sonora	<a href="#">@PespSonora</a>	Estatal
Policía Federal Nuevo León	<a href="#">@PFNuevoLeon</a>	Estatal
PGJE Guanajuato	<a href="#">@PGJEGUANAJUATO</a>	Estatal
PGR Aguascalientes	<a href="#">@PGR_Ags</a>	Estatal
PGR Baja California	<a href="#">@PGR_BC</a>	Estatal
PGR Baja California Sur	<a href="#">@PGR_BCSur</a>	Estatal
PGR Campeche	<a href="#">@PGR_Camp</a>	Estatal

Nombre	Twitter	Ámbito
PGR Chihuahua	@PGR_Chih	Estatal
PGR Chiapas	@PGR_Chis	Estatal
PGR Coahuila	@PGR_Coah	Estatal
PGR Durango	@PGR_Durango	Estatal
PGR Estado de México	@PGR_mx	Estatal
PGR Guerrero	@PGR_Guerrero	Estatal
PGR Hidalgo	@PGR_Hgo	Estatal
PGR Morelos	@PGR_Mor	Estatal
PGR Nuevo León	@PGR_NL	Estatal
PGR Querétaro	@PGR_Qro	Estatal
PGR Quintana Roo	@PGR_QRoo	Estatal
PGR Sinaloa	@PGR_Sin	Estatal
PGR Sonora	@PGR_Sonora	Estatal
PGR Tabasco	@PGR_Tab	Estatal
PGR Tamaulipas	@PGR_Tamps	Estatal
PGR Veracruz	@PGR_Ver	Estatal
PGR Yucatán	@PGR_Yuc	Estatal
Policía Estatal Querétaro	@poesqro	Estatal
DSPM Chihuahua	@policiachih	Estatal
Prevención del Delito SSP Puebla	@PrevencionSSP	Estatal
Seguridad Guanajuato	@SeguridadGto	Estatal
Secretaría Seguridad Estado de México	@SS_Edomex	Estatal
SSP CDMX	@SSP_CDMX	Estatal
SSP San Luis Potosí	@SSP_SLP	Estatal
SSP Sonora	@ssp_sonora	Estatal
Unidad de Contacto del Secretario SSP CDMX	@UCS_CDMX	Estatal
Fiscalía de Chihuahua	@Fiscalia_Chih	Estatal
PGJ CDMX	@PGJDF_CDMX	Estatal
Comisión Nacional de Seguridad	@CNSeguridadmx	Federal
PGR	@PGR_AIC	Federal
Policía Federal	@PoliciaFedMx	Federal
SEDENA	@SEDENAmx	Federal
Emergencias 911 Chetumal	@911chetumal	Municipal
Rescate 911 San Pedro, Nuevo León	@911SanPedro	Municipal
C4 Xalapa	@c4_ver	Municipal
C4 Naucalpan	@C4Nauc	Municipal
FSPE Guanajuato	@FSPE_Gto	Municipal
Policía Municipal Mexicali	@MexicaliDspm	Municipal
Policía Municipal Hermosillo, Sonora	@Policia_HMO	Municipal
Policía Municipal Piedras Negras, Coahuila	@Policia_PNegras	Municipal
Policía Guadalajara, Jalisco	@PoliciaGDL	Municipal
Policía Mazatlán	@policiamazatlan	Municipal
Policía Zapopan	@PoliciaZapopan	Municipal
Policía Municipal San Luis Potosí, San Luis Potosí	@SPM_SLP	Municipal

Fuente: elaboración propia, 2019

Tabla 2. Listado de palabras clave utilizados.

Abuso	Asesinada	Delitos	Justicia	Prevención del delito
Abuso sexual	Asesinadas	Delitos Cibernéticos	Ladrón	Puñaladas
Acosarle	Asesinado	Delitos de alto impacto	Ladrones	Quemarropa
Acoso	Asesinato	Denuncia	Lesionados	Rematar
Acoso Sexual	Asesino	Denunciaron	Lesiones	Rematarlo
Agresión	Asesinos	Denuncias	Levantado	Riña
Ajuste de cuentas	Atentado	Detención	Linchado	Robaron
Alerta de Género	Atracarlo	Detenido	Linchamiento	Robo
Allanamiento	Atraco	Disparan	Macabro	Robó
Altercado	Bala	Disparo	Mafia	Robos
Amagar	Balacera	Droga	Magistrado	Sanguinarios

Amagó	Balean	Ejecutados	Mano Armada	Secuestrada
Amenaza	Bandido	Enfrentamiento	Masacre	Secuestrado
Amenazada	Bandidos	Explosión	Muere	Secuestro
Amenazas	Cadáver	Extorsión	Muertos	Seguridad
Amenazó	Captura	Extorsiones	Multihomicidio	Sicarios
Aprehendimos	Capturados	Fallece	Murió	Sospechoso
Aprehensión	Cárcel	Fallecido	Narco	Terrorismo
Arma	Cárteles	Fallecidos	Narcofosas	Tiro
Arma de Fuego	Castigo	Falleció	Narcomanta	Tortura
Armado	Chantaje	Feminicidio	Narcotráfico	Torturada
Armas	Ciberbullying	Feminicidios	Operativo	Torturadores
Arrojada	Condena	Flagrancia	Pandilla	Ultrajada
Arrojan	Corrupción	Guardia Nacional	Pedofilia	Víctima
Asaltantes	Crimen	Herido	Peleando	Víctimas
Asaltar	Cuchillo	Homicidio	Persecución	Vigilancia
Asaltaron	Decapitados	Huachicoleo	Pistola	Violación
Asalto	Defensa de las mujeres	Huachicolero	Plan de Seguridad	Violaciones
Asaltó	Delincuente	Inseguridad	Presos	Violencia
Asaltos	Delito	Jurisdiccional	Presos Políticos	Violento

Fuente: elaboración propia, 2019

- **Validación de palabras clave:** se lleva a cabo mediante el uso de la herramienta de TweetReach, con ella se permite generar reportes de Twitter Analytics (Union Metrics, 2020). Estos reportes determinan la importancia y relevancia de una palabra clave, con base en los indicadores de “Alcance Estimado” y “Exposición (Interacciones)” de un tweet, a partir de informes de tomas instantáneas de Twitter (Twitter Snapshot Reports), los cuales recuperan una muestra aleatoria de 100 tweets de los últimos días en donde mencionen la palabra clave buscada.

A partir de los reportes generados por cada uno de los principales términos identificados se seleccionaron los indicadores de “alcance estimado” (*estimated reach*) y “exposición del tweet” (*exposure*); de estos indicadores emana la variable de “pertinencia de la palabra clave”, con la cual se valida la interacción que tienen los términos seleccionados dentro de la red social online de Twitter (ver Tabla 3).

Tabla 3. Validación de las palabras clave.

No.	Palabra Clave	Alcance Estimado	Exposición (Interacciones)	Pertinencia de la Palabra Clave
1	Prevención del delito	1,074,307	3,838,922	72.02%
2	Aprehendimos	1,514,997	5,242,496	71.10%
3	Amenazada	729,112	2,226,465	67.25%
4	Amagó	738,810	1,607,486	54.04%
5	Sicarios	2,507,355	5,328,253	52.94%
6	Atraco	2,838,436	6,019,241	52.84%
7	Denunciaron	3,354,910	7,071,119	52.55%
8	Corrupción	886,412	1,535,690	42.28%
9	Atracarlo	2,146,801	3,651,734	41.21%
10	Huachicolero	117,113	194,986	39.94%
11	Balean	2,698,616	4,128,677	34.64%
12	Flagrancia	1,195,124	1,797,536	33.51%
13	Fallecidos	564,291	831,395	32.13%
14	Narcofosas	64,038	91,502	30.01%
15	Delitos Cibernéticos	126,875	177,050	28.34%
16	Narcomanta	1,391,236	1,900,877	26.81%
17	Mafia	50,418	67,358	25.15%

No.	Palabra Clave	Alcance Estimado	Exposición (Interacciones)	Pertinencia de la Palabra Clave
18	Defensa de las mujeres	221,692	296,036	25.11%
19	Acoso Sexual	2,110,298	2,808,787	24.87%
20	Puñaladas	375,286	497,175	24.52%
21	Ciberbullying	35,039	45,789	23.48%
22	Balacera	8,825,898	11,375,150	22.41%
23	Pedofilia	57,632	73,995	22.11%
24	Abuso	248,391	316,435	21.50%
25	Víctima	165,124	205,211	19.53%
26	Multihomicidio	10,646,744	13,195,508	19.32%
27	Abuso sexual	3,200,267	3,903,544	18.02%
28	Plan de Seguridad	136,400	165,548	17.61%
29	Narcotráfico	119,141	144,109	17.33%
30	Ajuste de cuentas	250,352	298,971	16.26%
31	Cuchillo	180,922	215,196	15.93%
32	Decapitados	131,961	156,942	15.92%
33	Ladrones	143,356	170,395	15.87%
34	Violación	420,678	499,614	15.80%
35	Riña	27,982	33,211	15.74%
36	Altercado	5,288,707	6,261,654	15.54%
37	Capturados	8,273,553	9,747,766	15.12%
38	Ultrajada	81,096	94,918	14.56%
39	Bandidos	113,979	133,121	14.38%
40	Operativo	1,006,558	1,172,413	14.15%
41	Arrojada	181,138	210,831	14.08%
42	Terrorismo	528,249	614,130	13.98%
43	Atentado	2,825,782	3,284,124	13.96%
44	Alerta de Género	5,561,691	6,456,436	13.86%
45	Víctimas	473,277	549,427	13.86%
46	Lesiones	3,646,254	4,230,919	13.82%
47	Lesionados	1,910,449	2,216,127	13.79%
48	Pistola	97,398	112,896	13.73%
49	Aprehensión	4,756,244	5,512,465	13.72%
50	Persecución	535,427	620,511	13.71%
51	Bala	745,034	862,343	13.60%
52	Homicidio	1,393,696	1,612,959	13.59%
53	Huachicoleo	664,447	768,976	13.59%
54	Torturada	68,026	78,690	13.55%
55	Asesinos	241,207	278,499	13.39%
56	Vigilancia	1,752,506	2,023,409	13.39%
57	Robaron	59,051	68,135	13.33%
58	Acoso	602,792	695,203	13.29%
59	Torturadores	930,857	1,073,495	13.29%
60	Asaltantes	616,034	710,399	13.28%
61	Ejecutados	208,892	240,761	13.24%
62	Bandido	77,125	88,887	13.23%
63	Condena	66,302	76,397	13.21%
64	Cárteles	3,818,037	4,398,890	13.20%
65	Justicia	314,563	362,403	13.20%
66	Asesinada	685,368	789,347	13.17%
67	Asesino	1,750,496	2,015,723	13.16%
68	Captura	4,603,592	5,300,552	13.15%
69	Feminicidio	1,859,634	2,141,004	13.14%

No.	Palabra Clave	Alcance Estimado	Exposición (Interacciones)	Pertinencia de la Palabra Clave
70	Murió	95,914	110,411	13.13%
71	Delincuente	447,193	514,678	13.11%
72	Jurisdiccional	150,521	173,176	13.08%
73	Arma de Fuego	3,928,752	4,518,378	13.05%
74	Asesinado	1,761,971	2,026,518	13.05%
75	Asesinato	1,757,997	2,021,945	13.05%
76	Cadáver	7,087,071	8,150,456	13.05%
77	Armas	4,389,324	5,047,575	13.04%
78	Chantaje	42,456	48,823	13.04%
79	Linchado	44,931	51,668	13.04%
80	Guardia Nacional	734,266	844,304	13.03%
81	Presos Políticos	64,641	74,327	13.03%
82	Secuestrado	346,470	398,385	13.03%
83	Seguridad	358,506	412,160	13.02%
84	Cárcel	104,102	119,658	13.00%
85	Delito	120,832	138,885	13.00%
86	Ladrón	765,876	880,295	13.00%
87	Macabro	49,839	57,284	13.00%
88	Presos	55,917	64,263	12.99%
89	Secuestro	82,785	95,140	12.99%
90	Armado	516,141	593,126	12.98%
91	Droga	269,920	310,185	12.98%
92	Magistrado	66,316	76,204	12.98%
93	Muere	355,133	408,109	12.98%
94	Quemarropa	64,379	73,980	12.98%
95	Asalto	7,219,962	8,295,965	12.97%
96	Asaltó	7,219,236	8,295,132	12.97%
97	Feminicidios	560,866	644,420	12.97%
98	Pandilla	218,580	251,159	12.97%
99	Peleando	72,724	83,561	12.97%
100	Muertos	111,706	128,306	12.94%
101	Arma	841,841	966,715	12.92%
102	Narco	3,977,163	4,567,413	12.92%
103	Sanguinarios	138,530	159,089	12.92%
104	Amenazó	170,576	195,857	12.91%
105	Inseguridad	469,787	539,449	12.91%
106	Mano Armada	897,057	1,030,029	12.91%
107	Rematar	174,524	200,386	12.91%
108	Extorsiones	4,797,779	5,508,578	12.90%
109	Masacre	869,816	998,676	12.90%
110	Amagar	263,963	302,976	12.88%
111	Asaltar	240,847	276,464	12.88%
112	Asaltos	1,064,585	1,222,010	12.88%
113	Enfrentamiento	2,024,017	2,323,282	12.88%
114	Denuncia	3,519,303	4,038,952	12.87%
115	Linchamiento	307,730	353,166	12.87%
116	Rematarlo	400,496	459,670	12.87%
117	Tortura	307,774	353,217	12.87%
118	Extorsión	889,346	1,020,554	12.86%
119	Falleció	317,898	364,826	12.86%
120	Robos	455,277	522,496	12.86%
121	Tiro	930,693	1,067,985	12.86%



No.	Palabra Clave	Alcance Estimado	Exposición (Interacciones)	Pertinencia de la Palabra Clave
122	Amenaza	618,970	710,199	12.85%
123	Levantado	368,534	422,885	12.85%
124	Secuestrada	409,661	470,037	12.84%
125	Asaltaron	510,969	586,169	12.83%
126	Disparan	537,072	616,088	12.83%
127	Violento	501,193	574,964	12.83%
128	Castigo	9,856,716	11,305,694	12.82%
129	Explosión	2,757,710	3,163,141	12.82%
130	Fallecido	913,015	1,047,289	12.82%
131	Arrojan	1,754,984	2,012,715	12.81%
132	Asesinadas	1,119,989	1,284,512	12.81%
133	Crimen	659,667	756,592	12.81%
134	Allanamiento	2,860,163	3,279,931	12.80%
135	Robo	7,669,133	8,795,210	12.80%
136	Robó	7,673,107	8,799,765	12.80%
137	Violaciones	732,774	840,368	12.80%
138	Agresión	1,603,439	1,838,502	12.79%
139	Delitos de alto impacto	4,200,345	4,816,539	12.79%
140	Detención	3,725,393	4,271,232	12.78%
141	Amenazas	2,461,520	2,821,732	12.77%
142	Delitos	4,894,005	5,610,539	12.77%
143	Fallece	4,756,432	5,452,847	12.77%
144	Detenido	1,583,115	1,814,414	12.75%
145	Herido	2,304,296	2,640,145	12.72%
146	Disparado	6,331,514	7,253,094	12.71%
147	Sospechoso	3,210,994	3,677,988	12.70%
148	Denuncias	3,471,768	3,976,433	12.69%
149	Violencia	4,791,362	5,486,390	12.67%
150	Acosarle	16,860	18,929	10.93%

Fuente: Elaboración propia con datos de Twitter, 2019.

- **Delimitación de clústeres de palabras clave:** se realizó una simplificación de la realidad y la interacción que tienen las palabras clave entre ellas, implicando cinco clústeres: Acontecimiento, Armamento, Delincuencia, Seguridad, Género/Sexuales. De cada uno de ellos se tiene una subclasificación por categorías de palabras clave, tal y como se detalla en la Figura 1, asimismo, se muestra en el diagrama el número de palabras clave que se tiene en cada clúster y en cada categoría. Por ejemplo, el clúster de Delincuencia cuenta con 22 palabras clave, estando compuesto por dos categorías denominadas: Delincuencia Organizada con 10 palabras clave; y, Delincuentes con 12 palabras clave.
- **Configuración del clúster de servidores:** se configuró un clúster de servidores el cual consta de 4 nodos, de estos, 3 se configuraron de manera virtual en la plataforma de Google Cloud™, y 1 nodo de manera local. La Figura 2, muestra la conectividad de los servidores utilizados para hacer las consultas al API<sup>1</sup> de Twitter (2016), así como para almacenar en la base de datos los tweets que cumplan con la consulta y el visualizador de los datos.

<sup>1</sup> API: Interfaz de programación de aplicaciones (por sus siglas del inglés Application Programming Interface), es un conjunto de rutinas que provee acceso a un software determinado, en este caso, el API permite que Logstash se pueda conectar a la red social online de Twitter para descargar de manera autónoma los Tweets (Twitter, 2016).

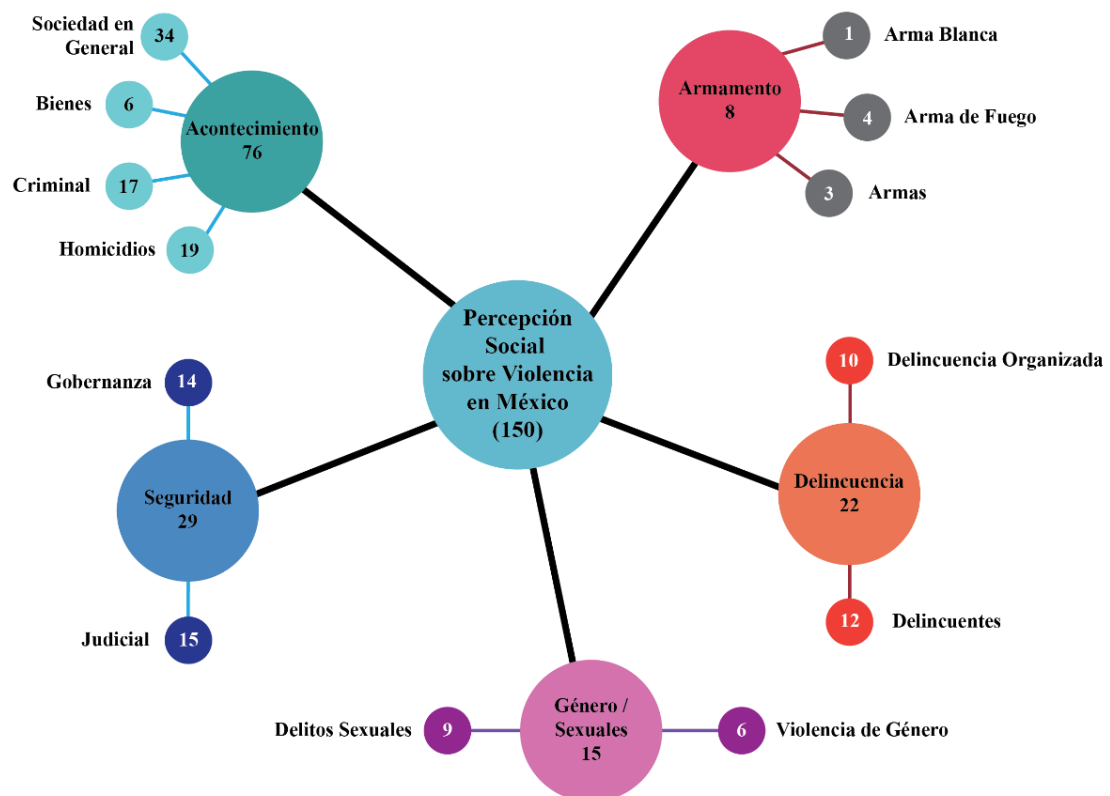


Figura 1. Modelado de Clústeres de palabras clave.

Fuente: elaboración propia, 2019.

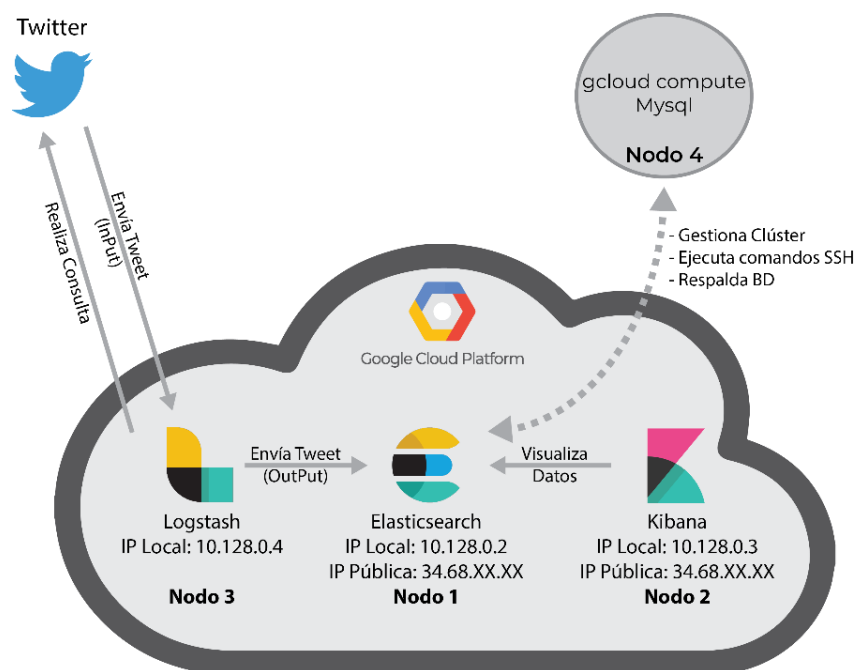


Figura 2. Configuración del clúster de servidores.

Fuente: elaboración propia, 2019.

- **Preprocesamiento de datos:** permite identificar aquellos tweets que cumplen con los requisitos deseados para el análisis de datos, así como estandarizar y normalizar la base de datos de tweets para que pueda ser comparable. Cada tweet es escrito por los usuarios y estos tienen sus propias particularidades de escritura. Utilizan caracteres especiales y emoticonos, aunado a que también restringen su vocabulario por la limitante del número de caracteres disponibles en Twitter. Esta fase consta de cinco pasos:
  - **Clasificación de los Tweets:** se tiene un universo de 20,736,887 tweets, los cuales fueron capturados del 14 de mayo al 27 de junio de 2019. En resumen, se capturaron tweets por 44 días, dando un promedio de 471,292 tweets diarios.
  - **Respaldo de la Base de Datos:** se generó mediante un archivo en formato JSON<sup>2</sup> con el respaldo de todos los tweets capturados. El respaldo creó un archivo de 103.3 Gb.
  - **Limpeza de datos:** excluye aquellos tweets mal capturados que proceden de un área geográfica externa a México (se da frecuente en las zonas fronterizas) y por nombres de ciudades que se repiten, sobre todo en América Latina; por el idioma, principalmente con el portugués por palabras semejantes y con el inglés ya que hay tweets que mezclan lenguaje. Además, se aplicaron 32 filtros a la base de datos para tener información georreferenciada a una escala de municipios.
  - **Normalización:** permite estandarizar los nombres de entidades federativas y municipios, corrección léxica de palabras, además de eliminar los símbolos y emoticonos.
  - **Validación:** se determina a partir de retomar la propuesta de utilizar el grado de veracidad sobre un conjunto de tweets adquiridos. Con base en Ashwin, et al. (2016), se logró medir el grado de veracidad (la Figura 3 muestra la representación gráfica de este indicador), el cual se encuentra integrado por:
    - Índice de Difusión (ID): identifica que tan rápido la información se esparce en Twitter.
    - Índice de Extensión Geográfica (IEG): determina la diseminación de la información hacia las entidades federativas y municipios.
    - Índice de Tweets Relevantes (ITR): comprueba el impacto de aquellos tweets que son únicos, por lo que detecta si existen tweets basura o spam en la base de datos.

---

<sup>2</sup> JSON: Notación de objetos de JavaScript (por sus siglas en inglés JavaScript Object Notation), es un formato de texto liviano para el intercambio de datos (ECMA-International, 2013).

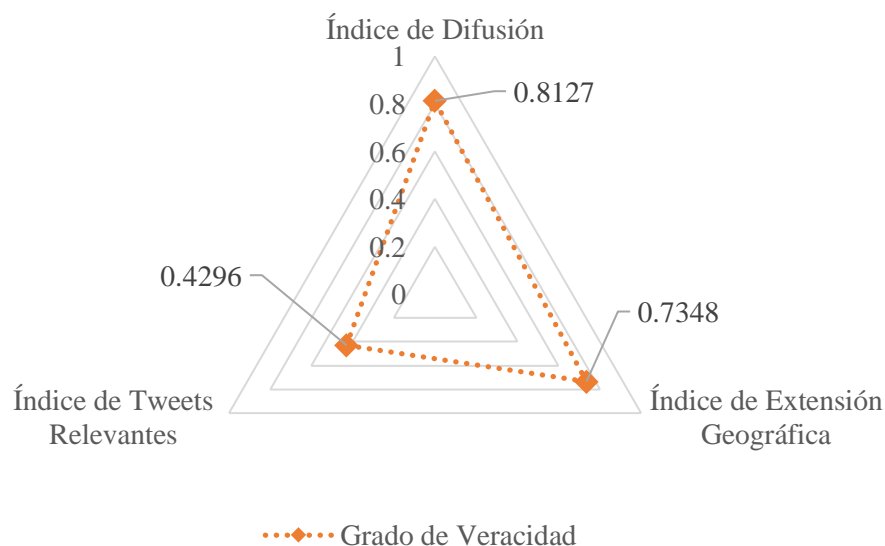


Figura 3. Grado de Veracidad de la base de datos.

Fuente: elaboración propia, 2019.

- **Procesamiento de datos:** corresponde a la manipulación de la información adquirida para mostrar la percepción social sobre la violencia en México. Para esto, se recurre a los siguientes niveles de análisis:
  - Detección de los principales hashtags,
  - Análisis por grafos.

## Resultados

### *Análisis por hashtags*

Al observar la Figura 4 se denotan los principales hashtags obtenidos desde la base de datos a partir de un análisis de nube de palabras (la cual fue desarrollada con la herramienta de Wordle). Los hashtags que aparecen con mayor frecuencia se representan con letras más grandes, mientras que los de menor frecuencia se representan con letras más pequeñas.

Al extraer los hashtags de los tweets almacenados en la base de datos se determina que existen 8,600 diferentes y únicos, los cuales son difundidos en 78,871 tweets y por 23,324 usuarios. La Figura 4, muestra los 50 principales hashtags, los cuales aglutinan a 24,773 tweets, representando 31.40% del total de tweets. Siendo “#DíaNaranja” la palabra clave más difundida, en donde de acuerdo con la Comisión Nacional para Prevenir y Erradicar la Violencia Contra las Mujeres (CONAVIM), el día naranja se conmemora el 25 de cada mes con el propósito de “actuar, generar conciencia y prevenir la violencia contra mujeres y niñas” (2019).



El grafo mostrado en la Figura 5, cuenta con 1,343 nodos y 152,316 aristas, representando a todos aquellos tweets que poseen una atracción entre ellos. Además, permite reconocer la existencia de hashtags o nodos con mayor relevancia e importancia, a partir del tamaño de los círculos, los cuales son proporcionales por el número de veces que fueron citados por los usuarios en sus tweets. Del mismo modo, las aristas se identifican por color y grosor de línea por su peso, esto es, la cantidad de veces que un nodo se relaciona con otro nodo. De este análisis se destacan las siguientes zonas: del lado izquierdo del grafo, se muestran hashtags orientados hacia los medios de comunicación; del lado inferior se ubican los hashtags vinculados con el tema de violencia, procuración e impartición de justicia; del lado derecho, información relativa a las demarcaciones de la geopolítica mexicana (estados o municipios); por último, en la zona superior, si bien no se destaca un nodo específico, pero si una gran cantidad de nodos interrelacionados, se encuentra la interacción de ámbito político, lo que llamaríamos la agenda pública del gobierno.

Retomando la clasificación y subclasificación por categorías de clústeres para la captación de los tweets ilustrada en la Figura 1, en donde se logró articular a los hashtags por clústeres, tal como se muestra en la Tabla 4, se enlazan los hashtags por correlación temática de acuerdo con la clasificación realizada para la selección de palabras clave utilizadas para la captación de los tweets.

Tabla 4. Resumen de clasificación de hashtags por clústeres de categorías.

Categoría	SubCategoría	Total Hashtags	Total Tweets Referencia
<b>Acontecimiento</b>	Bienes	3	60
	Criminal	26	586
	Homicidio	13	423
	Medios de Comunicación	332	16,919
	Migración	13	535
	Sociedad en General	374	17,692
<b>Armamento</b>	Arma de Fuego	1	25
	Armas	3	76
<b>Delincuencia</b>	Delincuencia Organizada	21	665
	Delincuentes	14	288
<b>Género/Sexuales</b>	Bullying	3	145
	Delitos Sexuales	17	582
	Discriminación Minorías	11	318
	Infantil	11	133
	Violencia de Género	41	3,467
<b>Seguridad</b>	Gobernanza	343	10,569
	Judicial	117	5,232

Fuente: Elaboración propia con datos de Twitter, 2019.

Los principales datos que resaltan son: del clúster de **Acontecimiento**, los temas de “sociedad en general” y “medios de comunicación”, en donde los usuarios de la red social online de Twitter expusieron que difunden información sobre diversos hechos de violencia que ocurren en su entorno; el clúster que sigue en importancia es sobre temas de **Seguridad**, en donde la “gobernanza” vinculada a acciones provenientes desde las instituciones de protección y procuración de justicia; y “judicial”, al seguimiento de las acciones legales ante acontecimientos que penetran en el tejido social, tales como asesinatos de personajes públicos, hechos de feminicidios, casos de corrupción, entre otros (ver Figura 6).



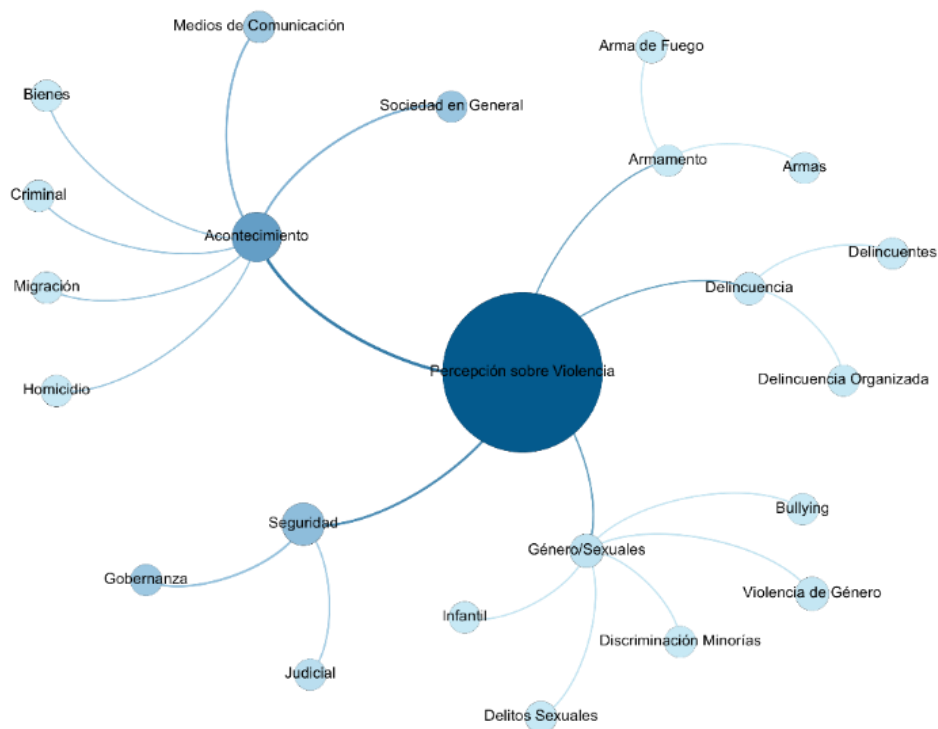


Figura 6. Grafo de clústeres de percepción sobre violencia en México.

Fuente: Elaboración propia utilizando la herramienta de Gephi, 2019.

## Conclusiones

Tomar como referencia a la red social online de Twitter como base para crear indicadores que permitan establecer el grado de violencia existente en un área geopolítica tiene sus riesgos. El primero de ellos está relacionado con la conectividad y el acceso a la red social, ya que las desigualdades económicas, tecnológicas, educativas, entre otras, pueden sesgar o dar un panorama parcial de lo que acontece en México, por lo que debe quedar claro que la información obtenida es una muestra poblacional de un sector de la población que utiliza esta red social online para transmitir información.

Otro riesgo detectado está vinculado con la metodología utilizada para homologar de forma léxica el contenido de los tweets. Como se observó durante la investigación, un tweet está limitado por el número de caracteres, por lo que los usuarios juntan palabras, quitan letras, se comunican con emoticonos, por mencionar algunas formas, con lo cual al analizar la información se le puede dar un contexto semántico incorrecto a lo que el mensaje quería expresar.

Este estudio mostro el potencial de utilizar la correlación de los tweets a través de un grafo para determinar la percepción social a partir de los temas de mayor impacto, en otras palabras, ilustra las palabras o frases que más se repiten en un periodo concreto. Además, este estudio introduce a un análisis semántico para la extracción de tweets vinculados con temas de violencia usando un corpus de palabras (banco de palabras). La metodología planteada permite que este estudio pueda ser aplicado en otras latitudes de América Latina, cambiando o adaptando el corpus de palabras para que sean coherentes con los modismos propios de cada región.

## Referencias

- Ashwin, K. T. K., Kammarpally, P., & George, K. M. (2016). Veracity of information in twitter data: A case study. *2016 International Conference on Big Data and Smart Computing, BigComp 2016*, 129–136. <https://doi.org/10.1109/BIGCOMP.2016.7425811>
- Bolgov, R., Filatova, O., & Semenova, E. (2017). Social media in Mexico, Argentina and Venezuela: Legal and political framework. *Proceedings of the 7th International Conference for E-Democracy and Open Government, CeDEM 2017*, 72, 253–259. <https://doi.org/10.1109/CeDEM.2017.33>
- Comisión Nacional para Prevenir y Erradicar la Violencia Contra las Mujeres. (2019). *¿Qué es el Día Naranja y por qué se conmemora?* <https://www.gob.mx/conavim/articulos/que-es-el-dia-naranja-y-por-que-se-conmemora>
- ECMA-International. (2013). *The JSON Data Interchange Format*. <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>
- Instituto Nacional de Estadística y Geografía. (2019). *Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2019*. <https://www.inegi.org.mx/programas/envipe/2019/>
- Ristea, A., Langford, C., & Leitner, M. (2017). Relationships between crime and Twitter activity around stadiums. *International Conference on Geoinformatics, 2017-Augus*, 5–9. <https://doi.org/10.1109/GEOINFORMATICS.2017.8090933>
- Sandoval-Almazan, R., & Gil-Garcia, J. R. (2013). Cyberactivism through social media: Twitter, YouTube, and the Mexican political movement “I’m Number 132.” *Proceedings of the Annual Hawaii International Conference on System Sciences*, 132, 1704–1713. <https://doi.org/10.1109/HICSS.2013.161>
- Sandoval-Almazan, R., Romero-Romero, A., Valle-Cruz, D., & Guadarrama-Chavez, A. (2014). Social media effect inside university communication: A Mexican case. *2014 International Conference on Education Technologies and Computers, ICETC 2014*, 110–114. <https://doi.org/10.1109/ICETC.2014.6998911>
- Santoro, E. (2014). Percepción Social. In *Psicología Social* (pp. 77–109). Trillas.
- Suárez-Serrato, P., Velázquez Richards, E. I., & Yazdani, M. (2018). Socialbots Supporting Human Rights. *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 290–296. <https://doi.org/10.1145/3278721.3278734>
- Twitter. (2016). *Tweets*. API Overview. <https://dev.twitter.com/overview/api/tweets>
- Union Metrics. (2020). *Social Analytics*. <https://tweetreach.com/social-analytics/>
- Vaidya, T., Votipka, D., Mazurek, M. L., & Sherr, M. (2019). Does Being Verified Make You More Credible? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. <https://doi.org/10.1145/3290605.3300755>