

Análisis de Redes Sociales Twitter

Data RoMa

22/04/2020

El análisis de redes sociales es una buena herramienta para conocer y entender la opinión pública usuaria de redes sociales. Cabe señalar que, dado el contexto socioeconómico en nuestro país y la brecha tecnológica característica de los países en desarrollo, este grupo recoge características específicas que merece un tratamiento diferenciado de los trabajos demoscópicos realizados en vivienda o por teléfono.

En México solo 70 por ciento de la población tiene acceso a internet (80 millones de mexicanos). Este segmento es el más urbano, más escolarizado y con mayores recursos económicos. Por ello, no es representativo de la opinión pública en su conjunto. Por su perfil es muy probable que este segmento represente solo a la clase media. Sin embargo estos segmentos pueden anticipar frecuentemente lo que será un tema de conversación social más general en el futuro.

Del total de usuarios de internet casi 90 % usa redes sociales; un poco más del 85 % usa WhatsApp, ligeramente por abajo está Facebook (84 %). Youtube lo utiliza sólo una tercera parte de los usuarios de Internet (31 %), Instagram ligeramente menos (27 %). Finalmente, Twitter sólo 10 % de los usuarios de internet.

Entre el público que accede a redes sociales la mayor parte de ellos dedican su tiempo a temas como deportes o espectáculos (alrededor de 85 %). Es solo una minoría los que emplean las redes sociales para debatir los temas de vida pública. Twitter es probablemente la red social con mayor frecuencia de opiniones en vida pública. Por ello, nuestro análisis se enfoca en esta red social.

El presente trabajo se realiza con interfaces basadas en la API de Twitter, donde se filtra- en tiempo real- los tweets de interés para crear una base de datos propia analizable. A partir de este análisis se identifican los topics, hashtags y utilizadores más importantes para el seguimiento. Este análisis se basa en un scrapping personal de tweets en México que hacen mención de COVID-19.

Tiempo

Los datos sobre tweets hasta 2020-04-20 08:08:25 tienen 161787 tweets.

El gráfico de tiempo presenta el número de tweets que mencionan el #COVID19MX. Estos tweets se muestran según la hora de su publicación desde el 30 de marzo al 20 de abril de 2020.

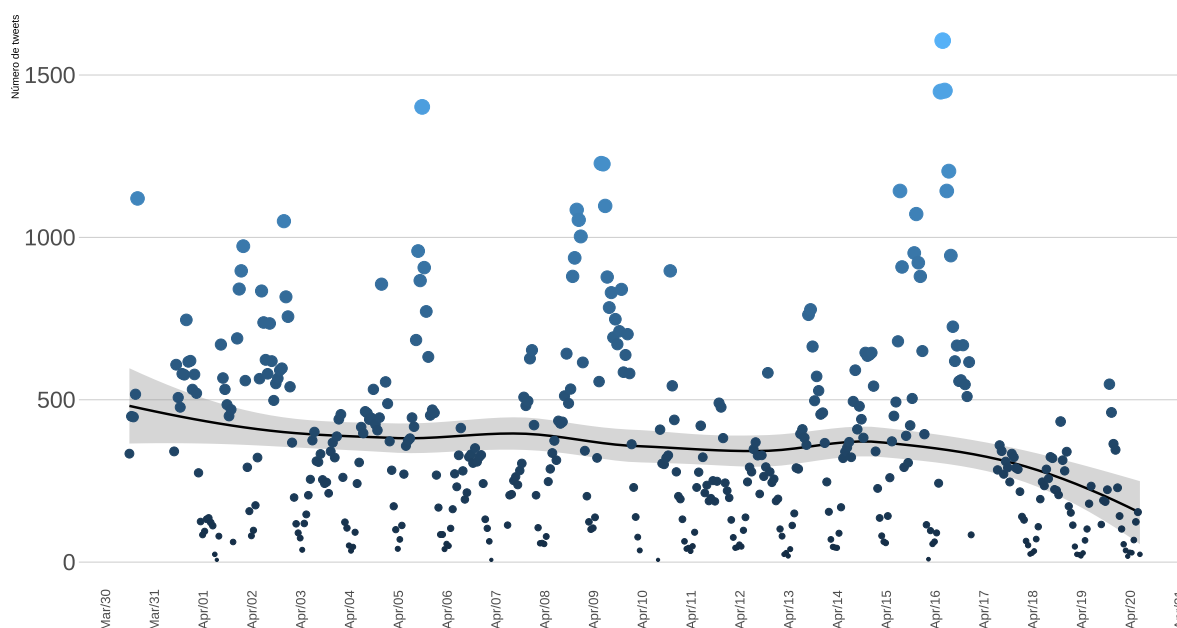
Es importante remarcar que los tweets tienen ciclos normales de creación determinado por las horas de descanso, así como la escalada esperada en los fines de semana. Este ejercicio permite evaluar la relevancia de un tema particular para la opinión pública usuaria de redes sociales y, subsecuentemente, entender su relación con eventos observados en un periodo de tiempo.

En la actualización de este reporte se incluyeron los tuits a partir del día 06 de abril de 2020. En este sentido, los puntos más altos fuera de tendencia correspondientes a este periodo se observan los días jueves 09 de abril y 16 de abril. El primero corresponde con el momento en el que se presentaron, por primera vez, las estimaciones de casos confirmados con base en el Modelo Centinela de Vigilancia Epidemiológica —estas cifras, debido a la carencia de una nota metodológica, causaron polémica e incluso confusión. El segundo coincide en fecha y horario con la conferencia matutina presidencial en la que el subsecretario Hugo López Gatell presentó un mapa desagregado a nivel municipal, cuyo propósito era ilustrar qué localidades, por un lado, terminarían la jornada de sana distancia el día 17 de mayo y cuáles, por otro lado, extenderían esta medida hasta al menos el 30 de mayo.

Otro elemento que vale la pena remarcar es que la tendencia respecto a este tema va a la baja: en general, se habla menos de forma cotidiana, salvo que exista un estímulo en forma de noticia o comunicados oficiales que lo impulse.

Frecuencia de tweets que mencionan al #COVID19mx

Tweets agrupados por hora del 31 de marzo al 20 de abril de 2020



Elaboración propia con datos de Twitter | @lolo7no | @guzmart_ | CEGAM-Data Roma

Figura 1: Línea de Tiempo de Tweets sobre COVID-19 en México - Twitter

[Link a esta gráfica](#)

En la actualización de este reporte se incluyeron los tuits a partir del día 06 de abril de 2020. En este sentido, los puntos más altos fuera de tendencia correspondientes a este periodo se observan los días jueves 09 de abril y 16 de abril. El primero corresponde con el momento en el que se presentaron, por primera vez, las estimaciones de casos confirmados con base en el Modelo Centinela de Vigilancia Epidemiológica —estas cifras, debido a la carencia de una nota metodológica, causaron polémica e incluso confusión. El segundo coincide en fecha y horario con la conferencia matutina presidencial en la que el subsecretario Hugo López Gatell presentó un mapa desagregado a nivel municipal, cuyo propósito era ilustrar qué localidades, por un lado, terminarían la jornada de sana distancia el día 17 de mayo y cuáles, por otro lado, extenderían esta medida hasta al menos el 30 de mayo.

Otro elemento que vale la pena remarcar es que la tendencia respecto a este tema va a la baja: en general, se habla menos de forma cotidiana, salvo que exista un estímulo en forma de noticia o comunicados oficiales que lo impulse. la República en el que dio a conocer su plan de acción económica para enfrentar la pandemia.

Hashtags, Menciones y Temas

Las nubes de palabras de hashtags, menciones y temas presentan las palabras y alusiones con mayor número de menciones en los tweets publicados del 30 de marzo al 20 de abril, relacionados al #COVID19MX. El tamaño de cada palabra en la nube alude a la mayor o menor frecuencia en que fue mencionada en las publicaciones.

Este ejercicio permite conocer los verbatismos más asociados con el tema de análisis y, con ello, analizar los constructos sociales que se crean en redes sociales respecto al mismo. Asimismo, esta información permite identificar aquellas figuras públicas o autoridades, así como movimientos detrás de los hashtags, a los que más alude la población usuaria de redes sociales con el fin de comunicar su interés particular en un tema.

Hashtags	Tweets	Temas	Tweets	Arrobas	Tweets
#QuedateEnCasa	21832	COVID	80006	@lopezobrador_	9641
#SusanaDistancia	10165	COVIDMX	62684	@hlgatell	8934
#Mexico	4698	COVIDMEXICO	36283	@ssalud_mx	7728
#CDMX	3465	SANA	22176	@tu_imss	3999
#ULTIMAHORA	2953	DISTANCIA	21358	@gobcdmx	3278
#QuedateEnCasaYa	2836	MX	21066	@m_ebrard	3196
#SNTEsalud	2678	CASOS	20886	@metrocdmx	2719
#Tlalnepantla	2502	MÉXICO	18019	@felipecalderon	2557
#SanaDistancia	2424	GOBIERNO	17262	@enriquealfaror	2180
#cuarentena	1986	CORONAVIRUS	14682	@gobiernomx	1932
#BuenMiercoles	1956	MEDIDAS	12199	@s	1684
#15Abril	1952	ASÍ	11785	@susanadistancia	1645
#CuandoTermineLaCuarentena	1949	CDMX	11785	@ssaludcdmx	1193
#QuedateEnLaCasa	1946	PANDEMIA	11550	@alynmon	1029
#AMLOseVA	1876	SI	11147	@drago237	887
#TeamBSide	1862	SALUD	10832	@a_vcarranza	764
#MexicoSolidario	1603	SUSANADISTANCIA	10750	@zoerobledo	735
#Feminicidios	1585	HOY	10557	@luisaalcalde	666
#EnEdomexPrevenirEsSalud	1521	ABRIL	10547	@alfredodelmazo	661
#LordMolecula	1427	CONFIRMADOS	9860	@gobiernozac	598



Figura 2: Menciones más comunes en Tweets sobre COVID-19 en México - Twitter

Para este periodo de tiempo, la discusión pública en Twitter es dominada por aquellos hashtags relacionados con las recomendaciones gubernamentales para contener el contagio del virus ([#QuedateEnCasa](#), [#SusanaDistancia](#)). Sin embargo, el primer hashtag ya dobla en número de menciones al segundo. Esto puede ser indicativo que el quedarse en casa, como estrategia pública para evitar la propagación del virus, retomó aún más importancia en días recientes. Por otro lado, [#Tlalnepantla](#) (tema surgido de un brote del virus en un hospital ubicado en este municipio) sigue como uno de los hashtags más mencionados; sin embargo, la desaceleración en sus menciones es evidente, pues no sumó más de 40 menciones entre el 06 y el 20 de abril. Por último, la Ciudad de México dobló el número de menciones respecto al periodo pasado; la capital del país es la entidad con mayor número de casos confirmados.



7

Definición: Es una etiqueta de metadatos precedida de un carácter especial (normalmente una arroba @) con el fin de relacionar a un usuario específico en una conversación y notificarlo de la misma. Se usa en servicios web tales como Twitter, Telegram, FriendFeed, Facebook, Google+, Instagram, Weibo para establecer relación entre una conversación y un usuario o para marcar la respuesta hacia un mensaje específico de un usuario. Resulta interesante observar que la discusión pública en redes se relaciona fuertemente con las autoridades correspondientes. Para el periodo de tiempo analizado, el presidente se posiciona como primer lugar con un total de 9,641; en comparación con el reporte anterior, esto significó un incremento de más del 150 %. El vocero del gobierno frente a la pandemia, Hugo López Gatell aparece en segundo lugar con un total de 8,938 menciones; equivalente a un incremento del 204 % respecto al periodo anterior. También vale la pena destacar que el gobernador del estado de Jalisco, Enrique Alfaro, es el único mandatario local que aparece en el top 10 con un total de 2,180 menciones.



Link a esta gráfica

Análisis de Sentimiento

El gráfico de Análisis de sentimiento presenta un promedio de la orientación positiva o negativa de los tweets publicados del 30 de marzo al 20 de abril, relacionados al #COVID19MX.

El texto de cada tweet puede ser positivo una vez que menciona palabras con este carácter, tales como: bueno, recuperación, cura, entre otras.

Por el contrario, el texto de cada tweet puede ser negativo una vez que menciona palabras con este carácter, tales como: malo, muerte, tristeza, entre otros. Este sistema automatizado permite asignar un puntaje para cada palabra en los tweets de la base de datos, lo cual obtiene un puntaje general por tweet. Con estos puntajes se obtiene un promedio de puntuación de los tweets por hora, y subsecuentemente construir el “sentimiento promedio” del día.

El puntaje se posiciona en una escala de números reales $(-\infty, +\infty)$, donde los valores positivos significan un “sentimiento promedio positivo”, los números negativos significan un “sentimiento promedio negativo” y el número cero significa un “sentimiento promedio neutral”. Cada círculo en el gráfico representa un día y su tamaño representa el volumen de tweets encontrados.

Análisis de sentimiento en tweets de COVID19

Cada círculo representa un día; el tamaño del círculo indica la cantidad de tweets encontrados por día. Una puntuación mayor a cero representa un sentimiento promedio positivo; una menor, un negativo.

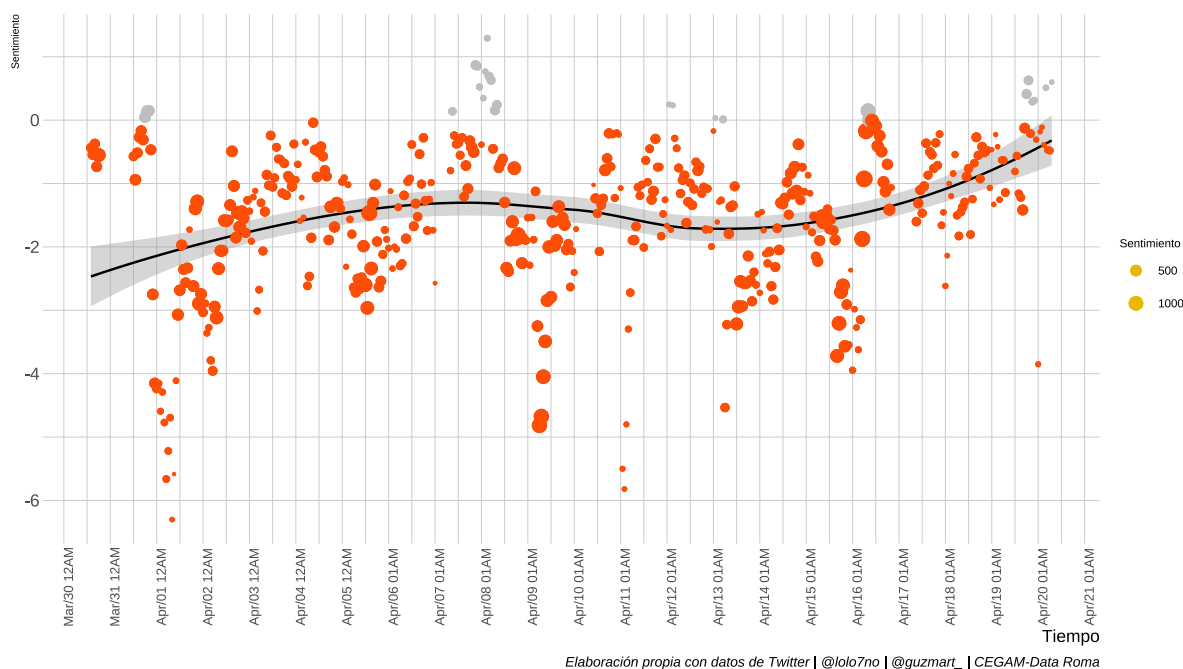


Figura 5: Análisis de Sentimiento en Tweets sobre COVID-19 en México - Twitter

[Link a esta gráfica](#)

Diez palabras positivas y 10 negativas

Ejemplo de Diccionario para Sentimiento

Negativas	Puntuación	Positivas	Puntuación
bastardo	-5.0	excepcional	5
negro	-5.0	magnífico	5
perro	-5.0	asombroso	4
pinchazo	-5.0	celestial	4
tragón	-5.0	eufórico	4
puto	-4.7	exuberante	4
burro	-4.0	fabuloso	4
catastrófico	-4.0	fantástico	4
culo	-4.0	ganador	4
defraudador	-4.0	gracioso	4

¡Emojis!

Debido a la propia extensión de un tweet, una forma no estructurada de expresión reside en las figuras que pueden acompañar el texto. Esta información se muestra en el gráfico de Emojis. Este gráfico presenta el número de veces en que el emoji específico fue usado en los tweets publicados del 30 de marzo al 20 de abril, relacionados al #COVID19MX. Este ejercicio es un primer acercamiento a entender la expresión no estructurada como un termómetro de la opinión pública en redes sociales en un tema específico.

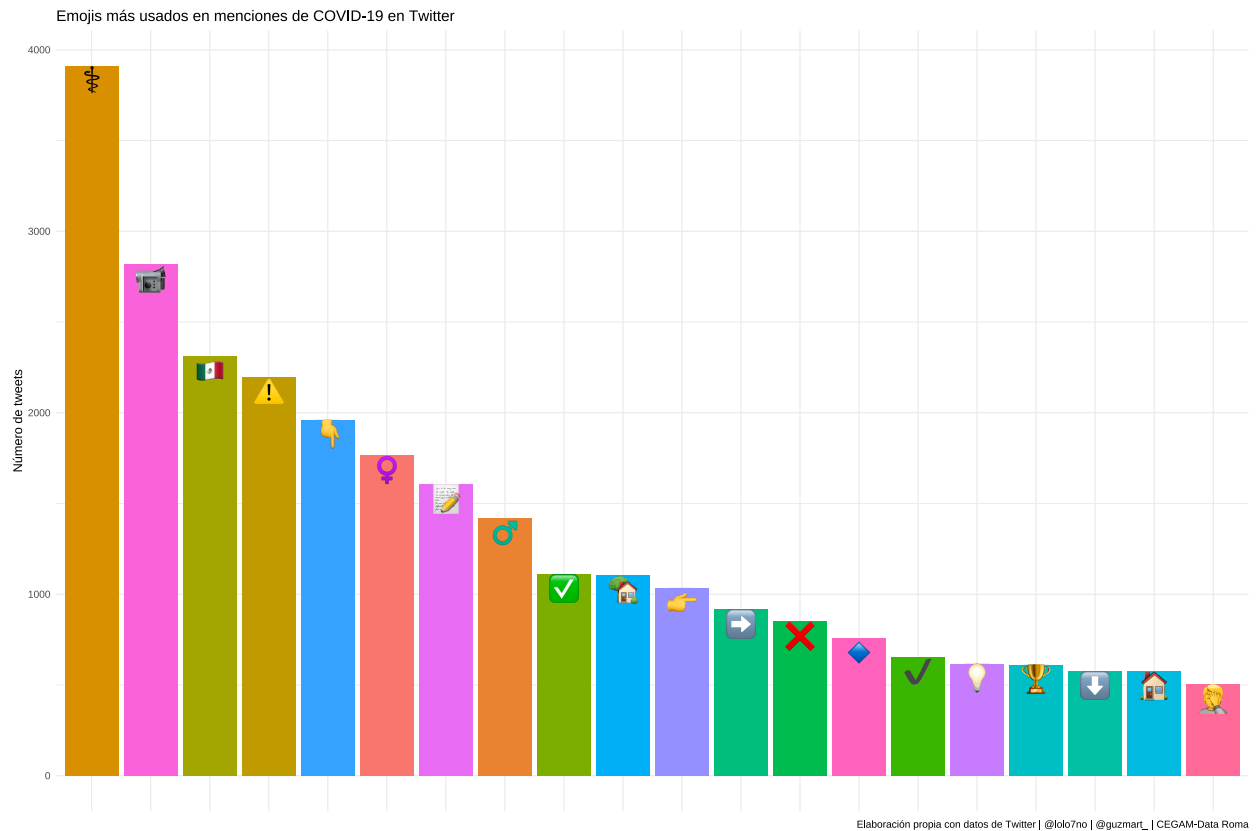


Figura 6: Emojis más comunes en Tweets sobre COVID-19 en México - Twitter

[Link a esta gráfica](#)