

Los inicios de la pandemia de COVID19 en Twitter. Análisis computacional de la conversación pública en lengua española

The beginnings of the COVID19 pandemic on Twitter. Computational Analysis of Public Conversation in Spanish Language

O início da pandemia COVID19 no Twitter. Análise computacional de conversas públicas na língua espanhola

Manuel Cebal-Loureda, Tecnológico de Monterrey, Monterrey, México
(manuel.cebral@tec.mx)

Gabriela Elisa Sued-Palmeiro, Investigadora Postdoctoral Conacyt Instituto de Investigaciones Sociales - Universidad Nacional Autónoma de México, Ciudad de México, México (gabriela.sued@sociales.unam.mx)

RESUMEN | Cuando comenzó la pandemia de COVID 19, las plataformas sociales tuvieron un rol central en la producción y acceso a la información. Este estudio identifica los tópicos de mayor interés y sus sentimientos asociados en Twitter en la conversación pública en lengua española en ese periodo. Asimismo, analiza el rol de Twitter como plataforma social involucrada en la conversación pública, como medio para la autocomunicación de masas y para amplificar la voz de un conjunto reducido de actores de alta visibilidad. Se recolectaron 231.375 tuits en España y América Latina durante dos meses. Se midieron indicadores de frecuencia y sentimientos, y se agruparon términos para identificar tópicos y determinar el interés de los usuarios sobre estos, mediante métodos digitales y de lenguaje computacional en R. La frecuencia de los principales términos es dinámica a lo largo del período estudiado, lo que sugiere diferentes percepciones de la pandemia. Los tópicos principales refieren a conversaciones en torno a la cantidad de casos, muertos y contagiados, con prevalencia de sentimientos negativos. La muestra analizada corresponde a mensajes generados por usuarios comunes en su gran mayoría, pero una parte de ella ha sido amplificada a gran escala mediante retuits y marcas de favoritos.

PALABRAS CLAVE: COVID-19; salud; Twitter; percepciones; métodos digitales; minería de textos; España; Latinoamérica.

FORMA DE CITAR

Cebal-Loureda, M. & Sued-Palmeiro, G. E. (2021). Análisis computacional de la conversación pública en lengua española. *Cuadernos.info*, (49), 1-25. <https://doi.org/10.7764/cdi.49.27467>

ABSTRACT | *At the beginning of the COVID19 pandemic, social platforms played a crucial role in the production and access to information. This study aims to identify the topics of most significant interest and their associated feelings during the onset of the pandemic on Spanish-language tuits. In addition, we analyzed the role of Twitter as a social platform involved in the public conversation, both as a means for mass self-communication and for amplifying the voice of a reduced set of high visibility actors. 231,375 tweets were collected in Spain and Latin America over two months. Then, the sample was analyzed with digital methods and techniques through computer programming in R. Frequency and sentiment indicators were measured, and terms were grouped to identify topics and determine users' interests. The frequency of the main terms is dynamic throughout the period studied, suggesting different perceptions of the pandemic. The main topics refer to conversations around the number of cases, deaths, and infections. Sentiment analysis shows the prevalence of negative feelings. The analyzed sample corresponds to ordinary users' messages for the great majority, but a part of it has been amplified on a large scale through retweets and bookmarks.*

KEYWORDS: COVID-19; health; Twitter; social media; digital methods; text mining; Spain; Latin America.

RESUMO | *No início da pandemia COVID-19, as plataformas sociais desempenharam um papel central na produção e no acesso à informação. Este estudo tem como objetivo identificar na conversa pública os temas de maior interesse e seus sentimentos associados no Twitter em língua espanhola nesse período. Além disso, foi analisado o papel do Twitter como plataforma social envolvida na conversação pública, tanto como meio de autocomunicação de massa quanto para amplificar a voz de um conjunto reduzido de atores altamente visíveis. Foram recoletados 231.375 tuites na Espanha e na América Latina, durante dois meses. A amostra foi analisada medindo indicadores de frequência e sentimentos, e os termos foram agrupados para identificar temas e determinar o interesse dos usuários sobre eles através de métodos digitais e da linguagem computacional R. A frequência dos termos principais é dinâmica durante o período estudado, sugerindo diferentes percepções sobre a pandemia. Os principais temas referem-se a conversas sobre o número de casos, mortos e contagiados. A análise mostra a prevalência de sentimentos negativos. A amostra analisada corresponde a mensagens geradas por usuários comuns em sua maior parte, mas parte dela foi ampliada em grande escala por meio de retuítes e marcações.*

PALAVRAS CHAVE: COVID-19; saúde; Twitter; mídia social; métodos digitais; mineração de texto; Espanha; América latina.

INTRODUCCIÓN

La pandemia COVID 19 puede considerarse un evento crítico e imprevisto, con consecuencias que trascienden el momento (Pantojas-García, n.d.), desde decisiones de salud pública hasta cambios en la producción económica que afectan el núcleo de la vida social e individual (UNDP, 2020). Su irrupción rápida en el contexto mundial, el desconocimiento general de su evolución y los cambios radicales que ha introducido en la vida cotidiana de millones de personas incrementaron el consumo y la circulación de información en la fase inicial de la pandemia (“¿Cómo es el consumo?”, 2020). Según el informe Reuters de consumo de noticias digitales, más de la mitad de los encuestados a nivel mundial accedió a noticias sobre la pandemia desde plataformas sociales (Newman et al., 2020).

Las redes sociales han sido consideradas por el periodismo y la consultoría como la puerta de entrada a noticias emitidas por los medios tradicionales acerca de la pandemia (Vega, 2020). Sin embargo, los usos de Twitter se extienden más allá de ser una red intermediaria entre prensa y audiencia, como un espacio para que los usuarios ordinarios expresen opiniones, ideas y sentimientos sobre asuntos públicos de importancia global (Rogers, 2019). Por su rol central en la construcción de conversaciones y tendencias de opinión pública (van Dijck, 2016) Twitter es, entre las plataformas sociales, la más indicada para estudiar cómo la pandemia COVID19 ha sido significada socialmente.

En un estudio previo (Sued Palmeiro & Cebal Loureda, 2020) se analizó el funcionamiento de las voces autorizadas en Twitter al inicio de la pandemia, definiéndolas como usuarios con perfil público con gran cantidad de seguidores. En este estudio se propone identificar los tópicos de mayor interés y sus sentimientos asociados en la conversación pública al inicio de la pandemia de COVID 19 en Twitter en Iberoamérica, así como determinar si los mensajes estudiados son representativos de la opinión pública de usuarios ordinarios que usan la plataforma para su autoexpresión, o bien reproducen pocos mensajes emitidos por usuarios de alta visibilidad, como medios de comunicación y líderes de gobierno y opinión, entre otros.

MÉTODOS DIGITALES Y ANÁLISIS COMPUTACIONAL DE TEXTOS APLICADOS A ESTUDIOS DE PLATAFORMA

Este trabajo se ubica en la interrelación de cuatro áreas de conocimiento: los métodos digitales (Rogers, 2019), el análisis computacional de textos (Castelfranchi, 2017), los estudios de plataforma (Plantin et al., 2018) y los de comunicación en salud (Rodríguez Roura et al., 2018).

Los métodos digitales se valen de los objetos digitales disponibles en la web, como métricas, geolocalizaciones y publicaciones, y los reorientan a la investigación

social mediante un conjunto de técnicas también digitales. Se plantean como una estrategia de investigación que estudia los cambios sociales y las condiciones culturales con datos en línea (Rogers, 2019). Esto no quiere decir que las plataformas digitales sean un reflejo de la vida social fuera de línea, pero sí que son indicadores de interés social y de las tendencias de la opinión pública (Rogers, 2018).

Los estudios de plataforma se conciben como un subcampo de los métodos digitales que, mediante la recolección de publicaciones, métricas y metadatos, estudian de forma crítica diferentes disposiciones, formas de uso y pautas culturales que distinguen a cada plataforma social (Burgess & Green, 2019).

En la última década, el estudio social de la web, centrado primero en sus usuarios y prácticas, se ha desplazado al análisis de datos en línea. Esta innovación requiere incorporar tecnologías digitales al proceso de investigación (Marres, 2017). Entre ellas, las técnicas de análisis computacional de textos, subcampo del procesamiento natural del lenguaje, permiten descubrir información nueva a partir de la extracción automática de patrones y tendencias en fuentes escritas (Moreno & Redondo, 2016).

Tanto los métodos digitales como los estudios de plataforma representan un desafío interesante para el denominado Sur Global (de Souza Santos, 2014), dado que existe un desequilibrio entre la producción centrada en datos de la literatura americana y europea y la producida en países iberoamericanos (Milan & Treré, 2017). Diferencias en idiomas, escuelas académicas de comunicación social, temáticas de interés y coyunturas respecto de los usos tecnológicos hacen que los métodos digitales y los estudios de plataforma necesiten una readaptación en la literatura en español, aportando nuevas temáticas, miradas y procedimientos al campo internacional (Meneses Rocha, 2018).

Este estudio adopta la estrategia de investigación de los métodos digitales para estudiar una condición sociocultural, la conversación pública y su significación social en los inicios de la pandemia de COVID 19, explotando datos extraídos de Twitter. Simultáneamente, intenta arrojar luz sobre pautas comunicativas en dicha plataforma, entendiendo que juega un rol específico en la predisposición a la producción y circulación de mensajes.

Twitter como fuente para el estudio de las expresiones públicas

Ubicada en decimocuarto puesto mundial de redes sociales más usadas¹, Twitter resulta relevante para el Sur Global, ya que el español es el tercer idioma más usado en la plataforma, luego del inglés y el japonés (“2018 Research on 100

1. <https://wearesocial.com/digital-2020>

Million Tweets”, 2018). Su estructura se centra en la comunicación de mensajes instantáneos y de rápida circulación sobre eventos relevantes del presente, entre una extensa comunidad internacional de usuarios que comparten intereses temáticos diversos (Bruns & Weller, 2016). Funciona como plataforma donde usuarios ordinarios comparten sus opiniones sobre diversos asuntos y como un medio para difundir mensajes producidos por usuarios de alta visibilidad (Kwak et al., 2010; Colleoni et al., 2014), definidos como usuarios que reciben muchos seguidores y logran gran cantidad de adhesión, por lo que sus mensajes tienen mayor posibilidad de propagarse (Omena et al., 2020). La conexión entre usuarios de alta visibilidad y ordinarios se hace posible mediante una serie de operaciones técnicas y métricas, tales como la cantidad de retuits y favoritos, o el uso de hashtags para seguir conversaciones públicas en tiempo real, sin barreras geográficas (Bruns & Highfield, 2016).

Los contenidos creados y distribuidos por usuarios de Twitter dejan trazas digitales que pueden ser recuperadas para la investigación social (Marres, 2017). La investigación con datos recolectados de Twitter ha contribuido a estudiar tendencias de opinión en la conversación pública y el debate político, así como a arrojar luz sobre su uso en contextos de crisis, entre otras temáticas de importancia (Burgess & Bruns, 2015).

Salud y COVID 19 en Twitter

En la última década, la comunicación en salud se concibe como un conjunto de procesos de interacción y construcción social de sentido a cargo de los diferentes actores implicados (Rodríguez Roura et al., 2018). El campo incorpora a sus investigaciones la recolección de datos producidos en redes sociales, el análisis supervisado de sentimientos y el análisis de redes (Arcila-Calderón et al., 2018). Durante la pandemia COVID 19, diversos trabajos apuntaron al rol de las plataformas sociales, especialmente Twitter. Una gran parte presenta métodos automatizados de recolección y análisis, destacando la referencia a temáticas recurrentes, palabras frecuentes, retuits y hashtags.

Existen precedentes respecto de la utilidad de Twitter como plataforma para comunicar novedades en el área de salud. Kullar y sus colegas (2020) se refieren al rol de Twitter como plataforma de circulación de mensajes sobre salud, tanto para profesionales médicos como para el público en general, en las pandemias de gripe aviar en 2009, ébola en 2014 y COVID 19 en 2020. La plataforma tiene valor para diseminar y acceder a las novedades científicas, así como para pacientes y sus familiares, quienes suelen agruparse para compartir información. Por otra parte, Rufai y Bunce (2020) analizan el uso de la plataforma por parte de los líderes políticos del G7 en el contexto de la pandemia. Su análisis de contenido reveló tres tipos de uso: el informativo,

que fue predominante, el de reforzamiento moral de la población, y la discusión política que busca plantear puntos de debate. En la misma línea, Haman (2020) revisa un corpus más extenso de expresiones en Twitter de líderes de gobierno, encontrando que 65% de los líderes de países pertenecientes a las Naciones Unidas han tuiteado en algún momento sobre la pandemia, incrementando su cantidad de seguidores.

Otra línea de trabajo, más cercana a la de este estudio, aborda la expresión de opiniones, percepciones y sentimientos mediante el análisis de mensajes, especialmente de las referencias léxicas. Han y sus colegas (2020) estudian la expresión de opiniones y sentimientos en las publicaciones en la red china Sina Weibo, el sitio más importante de microblogging en ese país, durante los estadios tempranos de COVID19, y clasifican los tópicos y subtópicos más referidos: recomendaciones de aislamiento social, bendiciones y plegarias, comentarios objetivos sobre la enfermedad, medidas de protección, voluntad de retorno al trabajo, entre otros. Los autores notan que la importancia de estos tópicos cambia según evolucionan los eventos sobre el coronavirus. Por su parte, Thelwall y Thelwall (2020) se enfocan en la perspectiva de género ante la pandemia, al comparar la frecuencia de palabras usadas por hombres y mujeres. Deportes y política son temáticas frecuentes entre los primeros, mientras que distancia social, casa y familia lo son entre las segundas. Por último, Thelwall y Levitt (2020) estudian las demandas y temáticas en torno a las personas con discapacidades en un conjunto reducido, pero ampliamente retuiteado, de publicaciones. Las principales temáticas son la necesidad de apoyo, las posibilidades de discriminación y las dificultades en el manejo de tecnologías.

Dos estudios relevantes abordan la relación entre las experiencias cotidianas de las pandemias y el encuadre de los medios sobre la enfermedad. Davis y sus colegas (2014) demostraron, con relación a la pandemia de A1HN1, que las percepciones cotidianas no pueden separarse de su marco mediático. Towers y colaboradores (2015) encontraron evidencia de que los medios de comunicación influenciaron la mayor parte de los tuits producidos durante la pandemia de ébola en África.

Otra importante línea de trabajo aborda la desinformación sobre la pandemia en plataformas sociales. En lo referente a Twitter, Pérez-Dasilva, Meso-Ayerdi y Mendiguren-Galdospín (2020) estudian los actores y las redes semánticas de desinformación mediante el análisis de redes. Los autores se refieren a la infodemia en el contexto de los debates políticos entre republicanos y conservadores en Estados Unidos, y entre críticos y favorables al gobierno en China. Destacan cómo, en un primer momento, se introducen falsos argumentos que exageran los peligros del coronavirus con vistas a agudizar la percepción de crisis económica –en el caso de Estados Unidos, ante la inminencia de sus elecciones presidenciales–, mientras que, en un segundo momento, las medidas de cuidado frente a la enfermedad adquieren mayor importancia.

Finalmente, en uno de los pocos trabajos que se enfocan desde el Sur Global, Gutiérrez y sus colegas (2020) estudian el uso de Twitter en Bahía Blanca, Argentina, mediante un análisis de redes sociales que les permite identificar que las conversaciones presentan un bajo grado de interacción y los actores más retuiteados no coinciden con su representatividad en la esfera pública. De acuerdo con estos trabajos, Twitter es un medio legítimo para distribuir información relevante sobre COVID 19, así como para que las audiencias compartan opiniones, sentimientos y percepciones.

MATERIALES Y MÉTODOS

Los datos fueron recolectados, procesados y visualizados a través del lenguaje de programación en R. Se usaron técnicas de minería de textos para descubrir nueva información (Moreno & Redondo, 2016). Las principales librerías utilizadas han sido *rtweet* (Kearney, 2020), *tidytext* (Robinson & Silge, 2020), *topicmodels* (Grün et al., 2020), *widyr* (Robinson, 2020), *ggraph* (Lin Pedersen, 2020) e *igraph* (Csardi & Nepusz, 2006).

La recolección se efectuó entre el 19 de marzo y el 16 de mayo de 2020, fase inicial de la pandemia. Este período permitió obtener las publicaciones del inicio del aislamiento social. Se recolectó un total de 231.375 tuits en español mediante la palabra clave coronavirus, accediendo a la Rest API (Application Program Interface) de Twitter con la librería *rtweet*. Se evitó recoger las réplicas de los retuits. Para captar cómo se ha ido asimilando la pandemia según pasaban las semanas y enfatizar la variable temporal, se efectuaron 13 recolecciones, 12 de ellas de 17.999 tuits y una de 15.392, con un intervalo de entre tres y cuatro días. El tamaño de las recolecciones responde al máximo permitido por Twitter para cada búsqueda con el método realizado.

Para cumplir con los objetivos propuestos, el procesamiento de datos tuvo dos orientaciones principales. Primero, se estudiaron los mensajes emitidos, en términos de frecuencia léxica, modelado por tópicos y análisis de sentimientos. En un segundo momento se recuperaron las métricas de reacciones de los mensajes, y estos se cuantificaron para determinar si corresponden a la expresión de opinión de un conjunto de usuarios ordinarios o, por el contrario, se amplifican los mensajes de pocos usuarios de alta visibilidad.

Para identificar las temáticas relevantes mediante análisis de frecuencia léxica, los tuits fueron desagregados en términos lingüísticos mínimos, proceso denominado *tokenización* e incluido en *tidytext*. Se aplicó la remoción de términos dependientes de contexto –como artículos, preposiciones y pronombres– a partir de la amplia lista proporcionada por Díaz (2016). Para evitar repeticiones no significativas,

se eliminó el término coronavirus, contenido en la totalidad de los mensajes. Se procedió del mismo modo con otros términos equivalentes. Así, los resultados de palabras frecuentes que se manejan en el artículo son aquellas que más veces aparecen junto a coronavirus, obviando este término y sus equivalentes directos. Utilizando como unidad mínima de cómputo estos términos lingüísticos o *tokens*, se efectuó la mayor parte de las operaciones de minería de textos, comenzando por el recuento de palabras (Robichaud & Blevins, 2011).

Complementariamente, se recuperó el tuit como unidad de significación para aplicar la técnica de modelado por tópicos, mediante el algoritmo Asignación Latente de Dirichlet (LDA), incluido en la librería *topicmodels*, también utilizado en el estudio de Ghan et al., (2020) ya reseñado. Este algoritmo clasifica todo documento como un conjunto de tópicos y cada tópico como un conjunto de términos (Blei et al., 2003). Con estas referencias, el algoritmo LDA crea un espacio vectorial no supervisado en el que va generando los tópicos según los vectores que relacionan términos y documentos.

LDA es un algoritmo complejo que admite diferentes configuraciones. Una de ellas es el valor *Alpha*, que hace que el algoritmo busque grupos más separados – para valores de *alpha* menores a uno– o bien grupos más combinados, en los que haya más términos comunes a varios tópicos –para valores de *alpha* mayores a uno–. En este caso, se configuró *alpha* con un valor menor que 1, específicamente 0,4, para que LDA genere tópicos más diversos. Se entiende que la variedad léxica en Twitter es bastante limitada y los temas están ya de por sí muy relacionados.

Otro de los valores configurables de LDA es *k*, que indica el número de tópicos generados. Este valor se ha establecido mediante un estudio previo para calcular la cantidad de tópicos más eficiente para un conjunto de datos. Se trata de la medida de perplejidad (Blei et al., 2003), disponible en la librería *topicmodels* mediante la función *perplexity* (Grün et al., 2020). Debido al extraordinario coste computacional de este cálculo, se ha efectuado sobre una submuestra de 10000 tuits. Los resultados para valores de *k* entre 2 y 11 han sido: 1970.461, 1791.831, 1732.825, 1725.598, 1760.093, 1781.241, 1817.099, 1851.048, 1894.675, 1943.602; siendo el más bajo el valor 1725.598, que corresponde a un valor de 5 para la variable *k*. De este modo, el estudio indica que el modelado por tópicos se optimiza para cinco tópicos. Cabe señalar que la numeración resultante de los tópicos es completamente aleatoria.

Por último, se aplicaron dos métodos distintos de análisis de sentimientos a las palabras más frecuentes del corpus: el clasificador Afinn (Nielsen, 2011) traducido al español (Mendoza Vega, 2018) y el NRC en su versión en español (Mohammad et al., 2013). El primero identifica si las interacciones contienen expresiones de estados de ánimo y si expresan sentimientos positivos o negativos.

El segundo asigna valores con respecto a varios sentimientos: miedo, confianza, tristeza, enfado, expectación, asco, alegría y sorpresa. Utilizando dos clasificadores, se obtienen perspectivas complementarias y se confirman tendencias detectadas.

Respetando las consideraciones éticas de AoIR (franzke et al., 2020), el estudio se basa en datos publicados en línea presentados de modo agregado, por lo que no produce perjuicios en terceros. Las bases de datos se anonimizaron respetando la privacidad de los usuarios.

RESULTADOS

En esta sección se presentan los hallazgos obtenidos mediante el descubrimiento automático de textos. Los tres primeros apartados responden al primer objetivo, ya que identifican los temas más relevantes de la conversación pública. El cuarto responde al segundo objetivo, dado que determina el rol de Twitter en la conversación pública sobre COVID 19 respecto de sus dos posibles usos: la amplificación de mensajes de usuarios altamente visibles o la expresión de opiniones de usuarios ordinarios.

Una narrativa alarmista: frecuencia léxica y su evolución

Las palabras más frecuentes de la muestra construyen una narrativa de alarma a partir de la prioridad que se le otorga a casos, muertos y pandemia. Esta se intensifica si se incluyen en este grupo palabras de significado cercano, como contagios, virus y fallecidos. Por el contrario, términos referidos a paliativos y medidas de salud, como cuarentena, salud, casa, médicos y medidas, tienen menor interés, tanto en frecuencia como en retuits. En tercer lugar, destaca otro agrupamiento que refiere al coronavirus en términos más estructurales y generales, donde pueden agruparse gobierno, país, crisis y mundo, y que percibe la pandemia como un hecho disruptivo que abarca espacios públicos y privados, afectando directamente a las personas y a la gente. Entre estas tres diferentes posibilidades –una enfermedad que pone fin a la vida, un conjunto de medidas de salud para contenerla, y una situación mundial imprevista que afecta a todas las personas y territorios–, la primera recibe más circulación y adhesión.

El gráfico 1 muestra la evolución temporal de la frecuencia de algunos términos. La frecuencia del término pandemia aumenta progresivamente desde mediados de marzo hasta el final de la muestra en mayo, obteniendo su pico a principios de ese mes. Junto a este término, casos y muertos también aumentan su frecuencia. Una evolución semejante tiene el término gobierno, aunque con una progresión más suavizada. Otros como salud o mundo siguen una evolución más irregular, aunque con un patrón semejante al que presentan casos y muertos: mayor frecuencia a principios y mediados de abril y al final de mayo. Por último, los términos crisis,

personas, hospitales, vida o médicos se mantienen constantes, lo cual no quiere decir que su presencia no sea importante.

Cabe destacar la evolución del término cuarentena. Durante los primeros días de recepción de COVID 19, la nueva situación de confinamiento parece ser lo que más afecta a la población. Se trata de las dos últimas semanas de marzo, momento en el cual se toman medidas de aislamiento social a nivel mundial y la vida de la gente cambia radicalmente en pocos días. Si bien comienza siendo el segundo término más frecuente en las primeras recogidas de datos, su presencia en la conversación pública desciende con el paso de las semanas: cae de más de mil apariciones en la recogida del 31 de marzo a la mitad, 504 apariciones, el 24 de abril.

Junto con la información que proporciona la frecuencia de palabras en los cuerpos de los tuits, los hashtags aglutinan la conversación pública en torno a un tema. De acuerdo con Wikström (2014), pueden clasificarse según funciones comunicativas. Hashtags como #cuarentena refieren a un tópico de interés que sucede más bien en el ámbito de la vida privada. En cambio, el hashtag #quedateencasa, también referente a un espacio privado, instala una acción colaborativa, empática y performativa que busca un cambio de hábitos y cuidados comunitarios. En el gráfico 2 puede observarse que es el más frecuente de esta muestra –siempre después de coronavirus y sus equivalentes directos–, manteniéndose en esa posición a lo largo del tiempo, aunque con cierto descenso desde principios de abril.

También desciende el hashtag #cuarentena, que se mantiene el segundo más frecuente en el conjunto. Por el contrario, el hashtag #pandemia sigue una tendencia ascendente, llegando a superar en frecuencia a #cuarentena, enfatizando una conversación diferente, ya no alrededor del cuidado sino de la preocupación. Ambas tendencias corroboran los hallazgos respecto de los términos más frecuentes, aunque en este caso las variaciones de frecuencia son más atenuadas, comprobando que el uso de hashtags se mantiene más constante.

Hay excepciones importantes a este comportamiento. En las capturas del 5 de mayo y del 30 de abril, surgen súbitamente dos hashtags ubicados en México relacionados con el riesgo y el hambre: #mexiquensesenriesgo y #méxicoconhambre. Por el volumen de sus publicaciones y su duración efímera, es posible conjeturar que obedecen a una estrategia local y concertada para insertar la temática de la enfermedad en un contexto de debate político. De hecho, estos hashtags coinciden en el tiempo con ascensos en la frecuencia de términos como gobierno en el gráfico 1. Ahora bien, también debe inferirse que, al mismo tiempo que la población se acostumbra al confinamiento y cae el debate acerca de la cuarentena, irrumpe con fuerza a finales de abril y principios de mayo la preocupación por la cuestión económica.

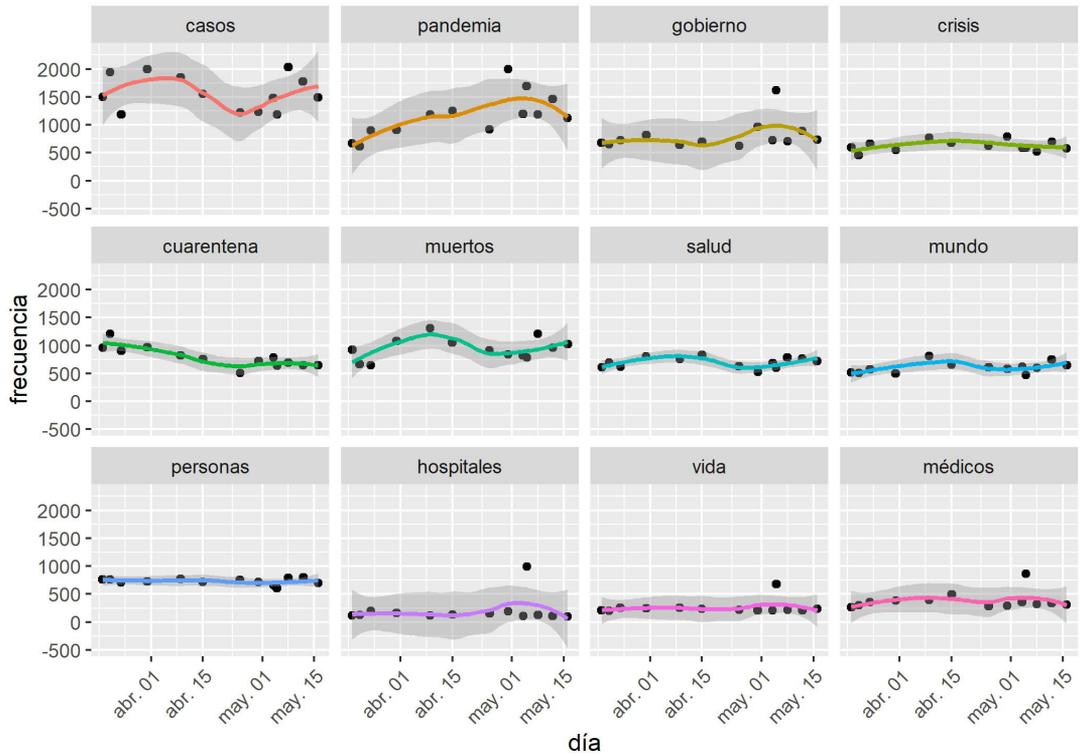


Gráfico 1. Evolución en el tiempo de una selección de palabras que con más frecuencia acompañan al término coronavirus

Fuente: Elaboración propia.

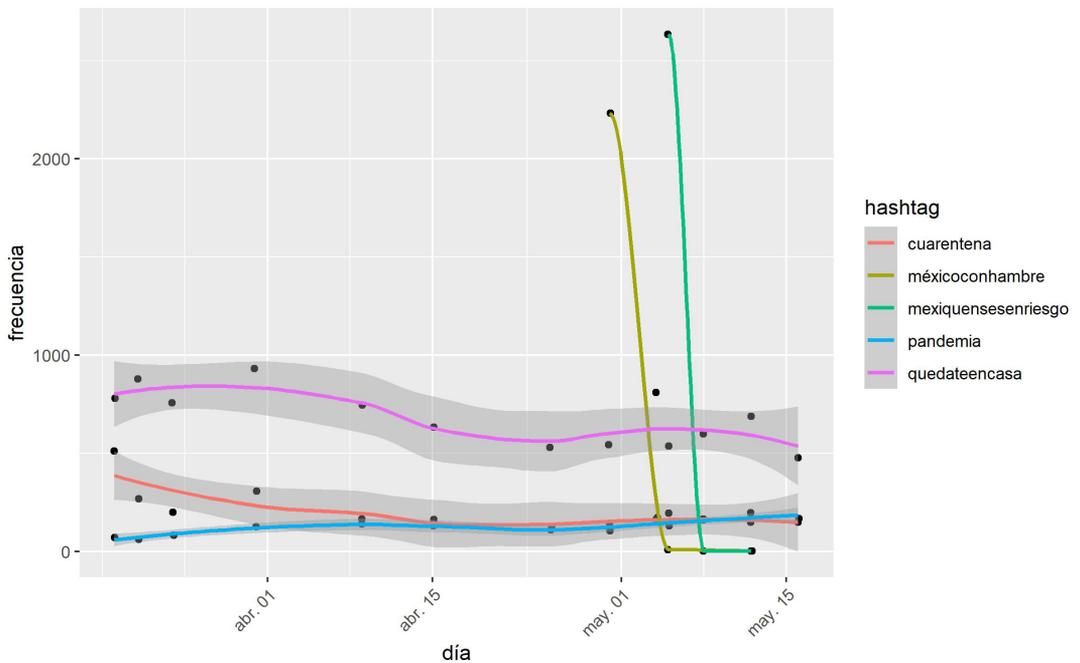


Gráfico 2. Detalle de la evolución de hashtags más utilizados en la muestra junto a la fuerte irrupción de dos hashtags localizados en México relacionados con la crisis

Fuente: Elaboración propia.

La evolución en la frecuencia de hashtags da cuenta de la variación en las conversaciones públicas: de una conversación más privada, centrada en la cuarentena y el cuidado, hacia otra más pública y pesimista, centrada en la pandemia y orientada hacia aspectos políticos y económicos. Esta tendencia confirma el análisis previo de frecuencia léxica.

Perspectivas de la pandemia: modelado por tópicos

Para conocer los temas de conversación generados respecto a COVID 19 en la muestra escogida, se procedió a un análisis sincrónico, mediante un modelado por tópicos. Para ello, se ha empleado el algoritmo LDA, introducido previamente, con los valores $\alpha=0,4$ y $k=5$.

Los resultados pueden consultarse en el gráfico 3, donde se visualizan en rojo los nodos correspondientes a cada tópico. Las palabras se vinculan con estos nodos según el valor β , que expresa el grado de pertenencia de un término a cada tópico. Este valor se ha enfatizado en el gráfico mediante vínculos de mayor grosor y opacidad. También se ha señalado la centralidad de cada palabra, esto es, la mayor cantidad de tópicos que conecta, mediante el tamaño de las etiquetas. La medida de centralidad es fundamental para leer correctamente la red de relaciones, ya que, como puede observarse, no siempre un mayor valor de β implica mayor centralidad. Por último, debido a las limitaciones espaciales y de legibilidad que supone este tipo de gráfico, es necesario indicar que solamente se muestran los términos que poseen un valor de β por encima de 0,0026.

Puede apreciarse que el tópico más caracterizado es el 4, al tener más términos por encima del umbral establecido para β en el gráfico y, en general, los términos aparecen más fuertemente vinculados. Este tópico estaría definido, principalmente, mediante términos como casos, muertos, pandemia, mundo o país. En conjunto, ofrece una perspectiva internacional o mundial del problema, ya que aparece la mayor parte de los países presentes en la conversación, incluida China, y entidades como la OMS. Muchos de los términos en este tópico se refieren a líderes políticos, nombres de países y términos numéricos, como cifra o número. Por ello, puede suponerse que está ligado a mensajes de los medios de comunicación, que producen precisamente el tópico más caracterizado de la muestra, a partir del cual se amplifica un relato basado en la cantidad de contagios y muertes.

El segundo tópico más caracterizado es el 3, cuyos términos principales son gobierno, pandemia, salud y crisis. Se reclama mayor acción frente a la situación económica. Posee también algunos términos que remiten a la respuesta colectiva y la solidaridad, como apoyo o enfrentar. El mensaje en conjunto se encuentra ligado a los dos hashtags que se instalan breve pero intensamente a principios de mayo, visibles en el gráfico 2. Esta conversación está centrada en México.

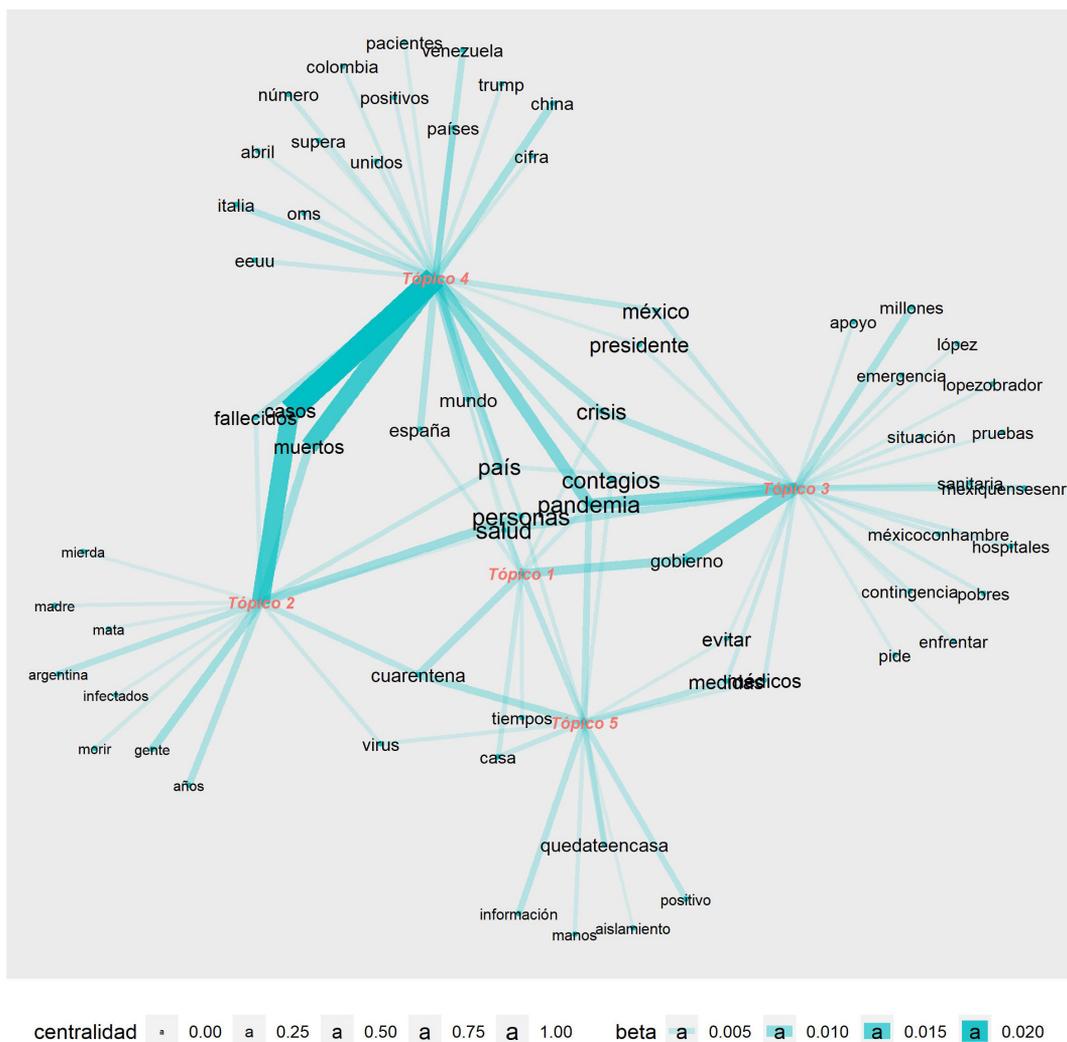


Gráfico 3. Modelado por tópicos mediante la función LDA de la librería topicmodels y gráficos mediante igraph y ggraph. Los valores de beta indican el grado de pertenencia de cada término al tópico graficado con el ancho y la opacidad de los vínculos. También se indica la centralidad de cada término en el conjunto de la red mediante el tamaño de la etiqueta

Fuente: Elaboración propia.

El siguiente tópico más caracterizado es el 2, a través de los términos casos, personas, gente, muertos, años y cuarentena. Semánticamente, posee un registro afectivo, refiriendo frustración y dolor por el alcance personal y mortal de la pandemia a gente de diferentes edades. Se trata de un tópico ubicado en Argentina.

El tópico 5, con menos palabras por encima del valor establecido para beta, estaría caracterizado por los términos cuarentena, pandemia, información, salud, positivo o medidas. Recoge preocupaciones transversales relacionadas con aspectos médicos y la forma de hacer frente al virus, mediante una conversación más racional y menos crispada, enfatizando medidas preventivas como lavarse

las manos y quedarse en casa para evitar contagios. Esta conversación no está ubicada en ningún país en específico.

Por último, el tópico 1 muestra una conversación basada en los términos gobierno, cuarentena, casa, pandemia y crisis. Si bien en este los términos definen con menos fuerza una temática propia, se aprecia una preocupación por la labor del gobierno y el presidente de cara a enfrentar la nueva situación. La conversación se localiza en España.

El algoritmo LDA detectó, además, valores de centralidad semántica respecto del conjunto de la conversación. Los dos términos con una mayor centralidad de la muestra son positivos: personas y salud. El término personas es además el más retuiteado de la muestra. A continuación, aparecen los términos pandemia, contagios, país y crisis.

Repercusiones afectivas: análisis de sentimientos

Además de los recuentos de palabras y su modelización en tópicos, se realizó un análisis de palabras mediante clasificadores de carga de sentimientos. El balance total del análisis es negativo: sumando todos los términos que puntúan negativamente en el clasificador Afinn, se alcanza un valor de -141253, frente a un valor de 66641 obtenido a través de la suma de todos los términos positivos. Distribuyendo esta carga a lo largo del eje temporal de la muestra, se encontró que este valor negativo es constante y apenas ofrece variaciones temporales.

En el gráfico 4 se detallan los diez términos que contribuyen a las puntuaciones positivas y negativas en el corpus estudiado. De este modo, es posible una perspectiva más semántica. Según el clasificador Afinn, los términos que más negativamente aportan son evitar, crisis y contagios, mientras que los positivos, con mucha menos repercusión total, son apoyo, ayuda y gracias.

Analizando semánticamente las palabras que más aportan a la puntuación Afinn en la muestra, puede observarse que la carga negativa tiene que ver con un conjunto de reacciones ante la pandemia, que ha traído lucha, dolor, hastío, frustración, miedo, hambre, riesgo y muerte. Todos estos sentimientos y sensaciones tienen prácticamente el doble de puntuación en la muestra que la semántica de las reacciones positivas, relacionada con el apoyo mutuo, la cooperación entre ciudadanos unidos, el agradecimiento a las instancias sanitarias y los anhelos de seguridad, libertad o amor que llegan a compartir los usuarios de la plataforma Twitter.

Esta carga negativa de la muestra se confirma a través del análisis mediante el clasificador NRC, gráfico 5. Si bien este clasificador es más específico a la hora de evaluar los sentimientos, los hay con connotaciones negativas frente a otros positivos. En este caso, predominan los valores altos en el sentimiento negativo de miedo, con una valoración por encima de 12.000 puntos. Le sigue el sentimiento

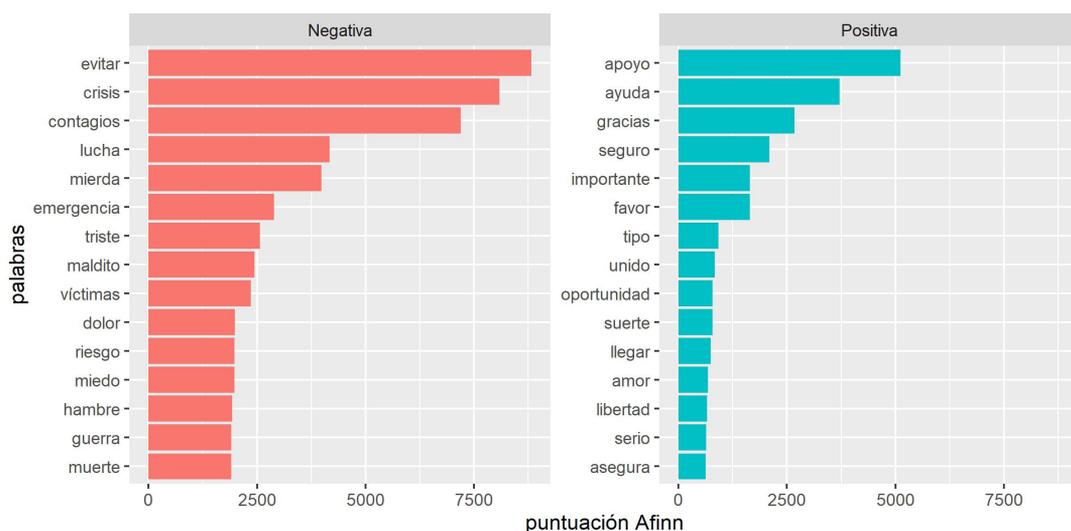


Gráfico 4. Palabras que más aportan en el corpus a los sentimientos negativos y positivos según el clasificador Afinn. En términos absolutos, el recuento de los términos negativos es mucho más alto

Fuente: Elaboración propia.

negativo de tristeza, con una valoración cercana a los 10.000 puntos. Sin embargo, con una valoración semejante, se encuentra la puntuación del sentimiento positivo de confianza, que también ronda los 10.000 puntos, resultando el sentimiento positivo predominante en la muestra. A continuación, aparecen los sentimientos de enfado y expectación –negativo y positivo, respectivamente– con una valoración ligeramente por encima de 6.000 puntos.

Si bien para el clasificador Afinn la carga negativa es predominante de manera constante a lo largo del periodo estudiado, es interesante observar la variabilidad en el tiempo de los resultados obtenidos mediante el clasificador NRC. Como se puede ver en el gráfico 5, la carga emocional en NRC baja ostensiblemente a mediados de abril, pero vuelve a subir, de manera aún más drástica, a finales de ese mismo mes, adquiriendo a principios de mayo los valores más altos, especialmente el sentimiento de miedo. Este momento coincide con el pico detectado en los hashtags #mexiquensesenriesgo y #méxicoconhambre del gráfico 2.

Con base en los análisis se puede afirmar que, más que repudio, enfado o tristeza, la pandemia COVID 19 ha provocado miedo en la población; en el otro extremo, hay muy poca puntuación para los sentimientos de sorpresa y alegría. Semánticamente, los términos de la muestra que más aportan a la alta puntuación del miedo son pandemia, cuarentena, evitar, lucha, hospital o emergencia, algunos de los cuales también producen altos valores de tristeza. Frente a ello, los términos de la muestra que aportan confianza son presidente, hospital -que aparece en el clasificador puntuando también como miedo-, personal, economía o importante.

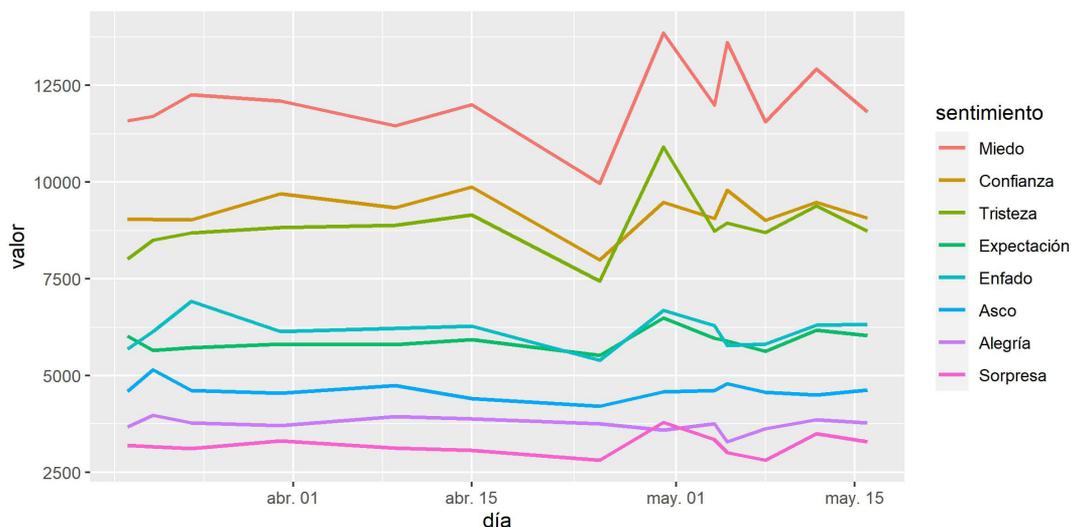


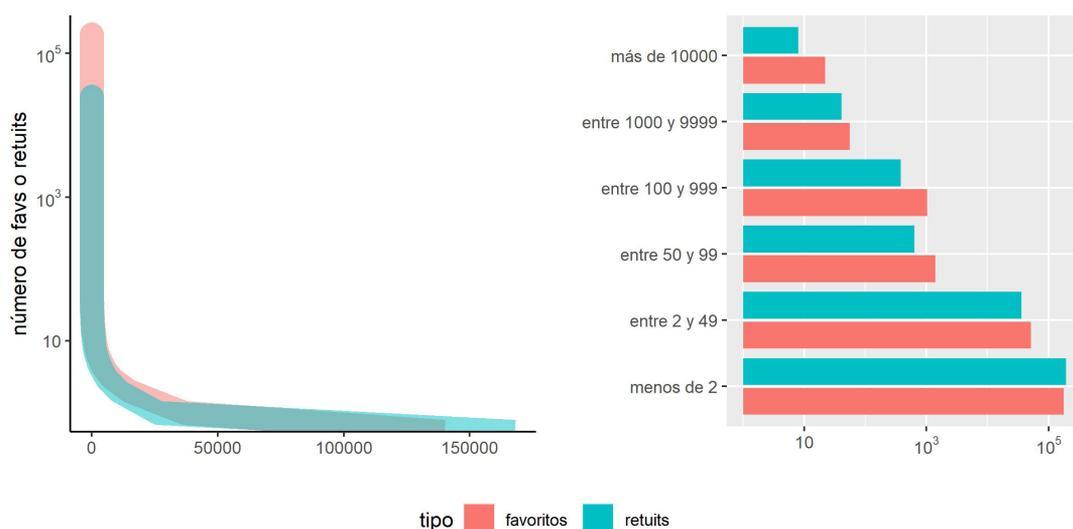
Gráfico 5. Evolución del sentimiento según el clasificador NRC en el análisis longitudinal a lo largo del periodo estudiado

Fuente: Elaboración propia.

Conversación y cámara de eco en torno a COVID19

Kwak, Lee, Park & Moon (2010) y Colleoni, Rozza y Arvidsson (2014) identifican dos funciones para Twitter: por un lado, se la considera una plataforma donde se desarrolla la conversación pública sostenida por audiencias participativas que crean, comparten y distribuyen contenidos, lo que Castells (2009) identifica como autocomunicación de masas; por otro lado, se la entiende como un potente medio para difundir noticias creadas por medios tradicionales y como instrumento de expresión para la política. En este último caso, los tuits funcionan amplificando y difundiendo a gran escala un conjunto reducido de voces públicas, a modo de cámara de eco. En la expresión sobre el coronavirus, las dos instancias se encuentran presentes.

Al trabajar con una muestra sin retuits, puede sostenerse que el análisis léxico representa suficientemente la expresión de los usuarios ordinarios, ya que contiene al menos un 99% de publicaciones que no se asumen como ampliaciones de usuarios de alta visibilidad. De este modo, la muestra tiene la forma de un gráfico de larga cola: una cabeza de pocas publicaciones muy amplificadas -izquierda del gráfico 6- y una cola de muchas publicaciones sin ampliación, que responden a la expresión de usuarios individuales. En la derecha del gráfico 6, puede observarse que los tuits de la muestra que recibieron más de mil retuits son menos de cincuenta, que los que recibieron entre cien y mil retuits son menos de quinientos, y que casi la totalidad de la muestra recibió menos de dos retuits.



Gráficos 6. Dos formas de mostrar la gran cantidad de tuits que tienen pocas interacciones y el gran número de interacciones que tienen muy pocos tuits

Fuente: Elaboración propia.

Por todo ello, la amplificación de voces resulta mayor que la autocomunicación. De hecho, si se atiende a la suma de retuits de los 49 tuits más replicados, prácticamente dobla al total de la muestra estudiada: solamente ocho publicaciones han sido replicadas más de diez mil veces. Sus voces se encuentran divididas entre el presidente de Argentina, el gobernador del estado de Jalisco, México, la presidenta de la comunidad de Madrid, dos periodistas populares de España y México, un youtuber, y dos medios españoles, *El país* y *ABC*, que han jugado un rol fundamental en la difusión de noticias sobre el coronavirus (Lázaro-Rodríguez & Herrera-Viedma, 2020).

El modelado de tópicos realizado previamente, especialmente la caracterización del tópico 3, junto al análisis de frecuencias, permite suponer que son los mensajes producidos por los medios los que impactan en la producción y circulación de mensajes entre usuarios ordinarios.

DISCUSIÓN Y CONCLUSIONES

Con relación al primer objetivo planteado –determinar los temas de mayor interés en la conversación pública–, se han identificado varios tópicos relevantes para la fase inicial de la pandemia. Si bien el más frecuentemente utilizado ha sido también el que recibe connotaciones más negativas, por su referencia a términos como pandemia, muertos, casos y contagios, el modelado por tópicos ha demostrado que es posible encontrar otras temáticas. Por ejemplo, el segundo

tópico más recurrente se refiere al cuidado de la salud, las medidas preventivas y la importancia de quedarse en casa. Este aspecto positivo se ve reforzado por el uso mayoritario del hashtag #quedateencasa, que es proactivo y que, frente a la diagnosis negativa que ofrecen las palabras frecuentes, propone de modo performativo la ejecución del único paliativo en la fase inicial de pandemia.

Mediante el modelado por tópicos se ha encontrado también que las palabras personas y salud son las más centrales, aun cuando no están entre las más frecuentes de la muestra. Esto indica que, a pesar de que la frecuencia de la conversación la marcan los casos, el gobierno, la pandemia y la cuarentena, semánticamente, lo que está en el centro de la conversación son las personas y la salud. Esta observación se refuerza atendiendo a la extraordinaria adhesión que el término personas obtiene en número de retuits. Por último, el modelado por tópicos también permitió identificar diferentes perspectivas sobre la pandemia según localizaciones geográficas: en México se destaca la crisis económica, en Argentina las medidas de gobierno, y en España, la cuarentena llevada a cabo en espacios privados.

El análisis de sentimientos arroja claramente un balance a favor de sentimientos negativos. Se observa que el miedo es el primero de la muestra, seguido por la confianza y la tristeza. Especialmente en la segunda mitad de abril, el sentimiento de tristeza supera en un pico al de confianza. Esto puede deberse a dos factores: por un lado, a mediados de abril se da el principal ascenso de muertes a nivel mundial²; además, los hashtags #mexicanosconhambre y #mexiquensesenriesgo -los únicos dos detectados como concertados en la muestra- tienen un rápido ascenso en la frecuencia de uso en el mismo período de fin de abril y principios de marzo. Puede conjeturarse que contribuyen, intencionadamente, a instalar el malestar entre los usuarios. Se necesita un estudio más profundo para corroborar esta hipótesis.

En una aproximación longitudinal, puede concluirse que la narrativa tiende a volverse más preocupante respecto de aspectos económicos y políticos, ya que el uso del hashtag #quedateencasa asume una curva descendente luego de las primeras semanas, coincidiendo con el uso de hashtags con tono de discusión política y queja -aun cuando puedan responder a una estrategia orquestada-, así como con el ascenso del hashtag #pandemia, que claramente connota un aumento de preocupación.

En general, las técnicas de recolección de datos y análisis de texto computacional permiten reorganizar la información en una lógica cronológica: fue posible

2. WHO COVID-19 Dashboard 2020- <https://covid19.who.int>

identificar la evolución de las percepciones sobre COVID 19 en sus inicios para una audiencia global de Twitter en español, e incluso se han avanzado percepciones acerca de una diferente recepción de la pandemia por países. Sin embargo, resta una discusión acerca de cómo los hallazgos producidos con datos en línea pueden ser vinculados con experiencias y datos fuera de línea (Rogers, 2019). Se entiende que, debido a la repercusión que actualmente tienen las plataformas, tanto en relación con su número de usuarios como con los estrechos vínculos que sostienen con otros medios de comunicación como la prensa, la televisión u otras plataformas sociales muy usadas como Facebook, son capaces de generar tendencias de opinión pública, instalar temas en la agenda pública y promover sentimientos en la población. El gran calado emocional que se genera en las redes sociales es, probablemente, una de las mejores muestras de que estos mensajes afectan de manera no meramente informativa.

En cuanto al segundo objetivo –determinar si prevalece en Twitter la amplificación de la palabra de usuarios de alta visibilidad o la autoexpresión de usuarios ordinarios–, el análisis de las interacciones revela que corresponden a una muestra formada en un 99% por tuits que no se asumen como retuits o réplicas exactas de voces autorizadas. Sin embargo, no hay que desestimar el aspecto de amplificación de usuarios de alta visibilidad, ya que la suma de los retuits de los 50 tuits más diseminados de la muestra supera en más del doble a los tuits con dos retuits o menos. Estos hallazgos contribuyen a develar que, para la muestra estudiada, la amplificación de mensajes producidos por usuarios de alta visibilidad es cuantitativamente más importante que la producción de mensajes por parte de usuarios ordinarios. Así, aunque resulta posible estudiar tendencias de opinión pública, se evidencia que la plataforma se ha usado en mayor medida para reproducir mensajes que para su producción.

La plataforma Twitter va más allá de ser un medio de reproducción de mensajes mediáticos a los que la opinión pública accede para su consulta, para ser un elemento central en la producción de interacciones que influyen en las opiniones públicas de usuarios ordinarios. Al observarse de manera masiva, indican tendencias generales de perspectivas y afectos sobre los acontecimientos que se construyen colectivamente dentro de la plataforma. Esta relación entre la amplificación de las voces públicas y la expresión de usuarios ordinarios, así como el análisis de contenido de las publicaciones más replicadas en el caso de COVID 19, han sido ya abordados en un estudio complementario que identifica cuáles son los actores de alta visibilidad relevantes y qué tópicos se expresan en ellos (Sued & Cebal , 2020).

Este estudio contiene algunas limitaciones. Primero, al recolectar datos con base en el idioma español se dejan de lado las particularidades que la pandemia

asume en cada país. Dado que las decisiones políticas y de salud pública frente a esta varían, se ven necesarios próximos estudios que desarrollen de manera más específica esta perspectiva. Segundo, el trabajo solo analiza los textos de Twitter; otros objetos digitales, como imágenes e hipervínculos, podrían ser relevantes. Tercero, una parte de la muestra puede responder a estrategias de posicionamiento de hashtags políticos; se necesita una investigación más profunda para saber de qué modo podrían afectar la generalidad de la muestra. Cuarto, los hallazgos obtenidos se validan para el universo de tuits recolectados en el período estudiado, la primera fase de la pandemia. Para profundizar en los hallazgos y comprender su dinamismo, sería necesario un monitoreo más extenso del tema.

Si bien es de suponer que en un momento de pandemia e incertidumbre las palabras frecuentes, tópicos y sentimientos tienden a la negatividad, a lo largo del artículo se han mostrado otras reacciones posibles, en concordancia con el estudio de Han y sus colegas (2020) sobre la red china Sina Weibo, que arroja una percepción distinta de la pandemia, centrada en tópicos como el control de la epidemia y la difusión de información del gobierno, opiniones sobre el cierre de las ciudades, medidas de protección de salud o la importancia del conocimiento antipandemia. Estas diferencias reafirman la relevancia de los condicionamientos culturales y geográficos a la hora de dar cuenta de las experiencias durante la pandemia y dimensionan todavía más los hallazgos de este artículo.

La bibliografía relevada (Davis et al., 2014; Towers et al., 2015) señala que los encuadres mediáticos se encuentran estrechamente relacionados con las percepciones cotidianas sobre la pandemia. Este hallazgo previo se confirma en el análisis del modelado por tópicos, especialmente la caracterización del tópico 3 que, junto al análisis de frecuencia, permite suponer que los mensajes de los medios impactan en la producción y circulación de mensajes entre usuarios ordinarios. Este nexo entre la amplificación de las voces públicas y la expresión de los usuarios, así como el análisis de contenido de las publicaciones más replicadas en el caso de COVID 19, deben abordarse en próximas investigaciones.

Este estudio también corrobora que, al menos en condiciones de crisis de salud, el sentido de la comunicación en salud se construye entre un conjunto heterogéneo de actores y expresiones (Rodríguez Roura et al., 2018). En este sentido, se amplían los hallazgos de Kullar y sus colegas (2020), quienes remarcan la utilidad de Twitter para estudiar la comunicación en salud, pero lo restringen a especialistas, profesionales y pacientes, así como los de Rufai y Bunce (2020) sobre su uso por líderes políticos, para diseminar información sobre cuidados. Este estudio añade una dimensión asociada a la opinión pública, menos especializada, pero más vinculada con las perspectivas acerca de la pandemia que expresan los usuarios no especializados.

Los hallazgos de este estudio resultan de interés para el área de la comunicación para la salud, al mostrar un tipo de conversación pública en el que convergen diferentes narrativas públicamente disponibles sobre la pandemia. Si bien el modelado por tópicos sugiere que la narrativa ligada a casos, muertes y contagios es la más promovida y la que genera más sentimientos negativos, otras narrativas de cuidados, prevención y atención médica están presentes. Este diagnóstico puede aprovecharse para insistir en la importancia de la presencia de líderes de opinión, instituciones de salud y autoridades en las redes sociales para difundir mensajes asociados con cuidados y prevención. Los hallazgos pueden considerarse para comprender cómo las redes sociales intervienen en la construcción colectiva de temáticas asociadas a la salud, en este caso, distribuyendo narrativas de corte pesimista por encima de propuestas de cuidado, prevención y actuación de los profesionales de la salud. Estudios posteriores podrán identificar qué actores sociales impulsan unas y otras, cuál es su nivel de aceptación por parte de los usuarios y, sobre todo, qué refuerzos son necesarios para usar las plataformas sociales como medio de difusión de cuidados y concientización en las medidas de salud para la población en general, antes que como medio para la difusión del miedo.

REFERENCIAS

- 2018 Research on 100 Million Tweets: What it Means for Your Social Media Strategy for Twitter. (2018, March 12). Vicinitas. Retrieved from <https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets>
- Arcila-Calderón, C., Alonso-Palacio, L. M., & García-Jiménez, A. (2018). Enfoques big data para la comunicación en salud: Análisis de redes y de sentimiento a gran escala (Big data approaches in health communication: Network analysis and large—scale supervised sentiment analysis). *Revista Científica Salud Uninorte*, 34(1). Retrieved from <http://rcientificas.uninorte.edu.co/index.php/salud/article/view/10544>
- Blei, D. M., Ng, Y. A., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3), 993-1022. Retrieved from <https://www.jmlr.org/papers/v3/blei03a.html>
- Bruns, A. & Highfield, T. (2016). Is Habermas on Twitter? Social Media and the Public Sphere. In A. Bruns, G. Enli, E. Skogerbø, A. O. Larsson, & C. Christensen (Eds.), *The Routledge Companion to Social Media and Politics* (pp. 56-73). New York, NY: Routledge.
- Bruns, A. & Weller, K. (2016). Twitter as a first draft of the present: And the challenges of preserving it for the future. In *Proceedings of the 8th ACM Conference on Web Science* (pp. 183-189). <https://doi.org/10.1145/2908131.2908174>
- Burgess, J. & Bruns, A. (2015). Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn. In G. Elmer, G. Langlois, & J. Redden (Eds.), *Compromised Data. From social Media to Big Data* (pp. 93-111). New York, NY: Bloomsbury Academic.

- Burgess, J. & Green, J. (2019). *YouTube: Online Video and Participatory Culture. Second Edition*. Cambridge, United Kingdom: Polity.
- Castelfranchi, Y. (2017). Computer-Aided Text Analysis: An Open-Aired Laboratory for Social Sciences. *Journal of Science Communication*, 16(2), C04. <https://doi.org/10.22323/2.16020304>
- Castells, M. (2009). *Comunicación y poder (Communication Power)*. Madrid, Spain: Alianza.
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily on Twitter Using Big Data: Political Homophily on Twitter. *Journal of Communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. In *InterJournal: Vol. Complex Systems* (p. 1695). Retrieved from <https://igraph.org>
- ¿Cómo es el consumo digital en México en el marco de la pandemia? (How is Digital Consumption in Mexico in the Context of the Pandemic?). (2020, May 5). *El Universal*. Retrieved from <https://www.eluniversal.com.mx/techbit/como-es-el-consumo-digital-en-mexico-en-el-marco-de-la-pandemia>
- Davis, M., Lohm, D., Flowers, P., Waller, E., & Stephenson, N. (2014). "We Became Sceptics": Fear and Media Hype in General Public Narrative on the Advent of Pandemic Influenza. *Sociological Inquiry*, 84(9), 499-518. <https://doi.org/10.1111/soin.12058>
- De Souza Santos, B. & Meneses, M. P. (2014). Prólogo (Foreword). In B. de Sousa Santos & M. P. Meneses (Eds.), *Epistemologías del Sur: Perspectivas (Epistemologies of the South: Perspectives)* (pp. 7-17). Madrid, Spain: Akal.
- Díaz, G. (2016). Stopwords-es. Retrieved from <https://github.com/stopwords-iso/stopwords-es/blob/master/stopwords-es.txt>
- franzke, a. s., Bechman, A., Zimmer, M., & Ess, C. (2020). *Internet Research: Ethical Guidelines 3.0*. Retrieved from <https://aoir.org/reports/ethics3.pdf>
- Grün, B., Hornik, K., & Blei, D. (2020). Topic Models (0.2-11). Retrieved from <https://cran.r-project.org/package=topicmodels>
- Gutiérrez, E. M., Viridis, J. M., Meller, L., Gurovich, C., Simón, F. D., & Leyes, L. F. (2020). ¿Retuitear o responder? Covid-19 y Twitter. El caso de la ciudad de Bahía Blanca - Argentina (Retweet or reply? Covid-19 and Twitter. The case of the city of Bahía Blanca, Argentina). *Awari*, 1(2). <https://doi.org/10.13140/RG.2.2.31471.87203>
- Haman, M. (2020). The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic. *Heliyon*, 6(11), e05540. <https://doi.org/10.1016/j.heliyon.2020.e05540>
- Han, X., Wang, J., Zhang, M., & Wang, X. (2020). Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *International Journal of Environmental Research and Public Health*, 17(8), 2788. <https://doi.org/10.3390/ijerph17082788>
- Kearney, M. W. (2020). rtweet: Collecting Twitter Data. (1.2.5). Retrieved from <https://cran.r-project.org/package=rtweet>

- Kullar, R., Goff, D. A., Gauthier, T. P., & Smith, T. C. (2020). To Tweet or Not to Tweet—A Review of the Viral Power of Twitter for Infectious Diseases. *Current Infectious Disease Reports*, 22, 14. <https://doi.org/10.1007/s11908-020-00723-0>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591–600). New York, NY: ACM. <https://doi.org/10.1145/1772690.1772751>
- Lázaro-Rodríguez, P. & Herrera-Viedma, E. (2020). Noticias sobre Covid-19 y 2019-nCoV en medios de comunicación de España: El papel de los medios digitales en tiempos de confinamiento (News about Covid-19 and 2019-nCoV in the Spanish Media: the Role of digital media in times of confinement). *El Profesional de la Información*, 29(3), e290302. <https://doi.org/10.3145/epi.2020.may.02>
- Lin Pedersen, T. (2020). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks (2.0.2). Retrieved from <https://cran.r-project.org/package=ggraph>
- Marres, N. (2017). *Digital sociology: The reinvention of social research*. Malden, MA: Polity.
- Mendoza Vega, J. B. (2018, April 15). Análisis de sentimientos con R - Léxico Afinn (Sentiment Analysis with R - Afinn Lexicon). *RPubs*. Retrieved from https://rpubs.com/jboscomendoza/analisis_sentimientos_lexico_afinn
- Meneses Rocha, M. E. (2018). Grandes datos, grandes desafíos para las ciencias sociales Big Data, Big challenges for social sciences). *Revista Mexicana de Sociología*, 80(2), 415-444. Retrieved from <http://revistamexicanadesociologia.unam.mx/index.php/rms/article/view/57723>
- Milan, S. & Treré, E. (2017, October 16). Big Data from the South: The beginning of a Conversation We Must Have. *DATACTIVE*. Retrieved from <https://data-activism.net/2017/10/bigdatasur/>
- Mohammad, S., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises* (arXiv:1308.6242). Retrieved from <https://arxiv.org/abs/1308.6242>
- Moreno, A. & Redondo, T. (2016). Text Analytics: The Convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(6), 57. <https://doi.org/10.9781/ijimai.2016.369>
- Newman, N., Fletcher, R., Schulz, A., Simge, A., & Kleis Nielsen, C. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf
- Nielsen, F. (2011). A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages* (arXiv:1103.2903). Retrieved from <https://arxiv.org/abs/1103.2903>
- Omena, J. J., Rabello, E. T., & Mintz, A. G. (2020). Digital Methods for Hashtag Engagement Research. *Social Media + Society*, 6(3), 2056305120940697. <https://doi.org/10.1177/2056305120940697>
- Pantojas García, E. (n.d.). La sociedad translocal: Notas para entender el cambio de época (The Translocal Society: Notes to Understand the Change of Age). CLACSO. Retrieved from <https://www.clacso.org/la-sociedad-translocal-notas-para-entender-el-cambio-de-epoca/>

- Pérez-Dasilva, J.-A., Meso-Ayerdi, K., & Mendiguren-Galdospín, T. (2020). Fake news y coronavirus: detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter (Fake news and coronavirus: Detecting key players and trends through analysis of *Twitter* conversations). *El Profesional de la Información*, 29(3), e290308. <https://doi.org/10.3145/epi.2020.may.08>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook. *New Media & Society*, 20(1), 293-310. <https://doi.org/10.1177/1461444816661553>
- Robichaud, A. & Blevins, C. (2011). Content-Based Analysis. *Tooling Up Digital Humanities*. Retrieved from http://toolingup.stanford.edu/?page_id=3
- Robinson, D. (2020). widyr: Widen, Process, then Re-Tidy Data (0.1.3). Retrieved from <https://cran.r-project.org/package=widyr>
- Robinson, D. & Silge, J. (2020). tidytext: Text Mining Using «dplyr», «ggplot2», and Other Tidy Tools (0.2.4). Retrieved from <https://cran.r-project.org/package=tidytext>
- Rodríguez Roura, S. C., Cabrera Reyes, L. de la C., & Calero Yera, E. (2018). La comunicación social en salud para la prevención de enfermedades en la comunidad (Social communication in health for disease prevention in the community). *Humanidades Médicas*, 18(2), 384-404. Retrieved from <http://www.humanidadesmedicas.sld.cu/index.php/hm/article/view/1302>
- Rogers, R. (2018). Otherwise Engaged: Social Media from Vanity Metrics to Critical Analytics. *International Journal of Communication*, 12, 23. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/6407>
- Rogers, R. (2019). *Doing Digital Methods*. London, United Kingdom: SAGE.
- Rufai, S. R. & Bunce, C. (2020). World Leaders' Usage of Twitter in Response to the COVID-19 Pandemic: A content analysis. *Journal of Public Health*, 42(3), 510-516. <https://doi.org/10.1093/pubmed/fdaa049>
- Sued Palmeiro, G. E. & Cebal Loureda, M. (2020). Voces autorizadas en Twitter durante la pandemia de COVID-19: Actores, léxico y sentimientos como marco interpretativo para usuarios ordinarios (Authorized Twitter voices during the COVID-19 pandemic: actors, vocabulary, and feelings as an interpretive framework for ordinary users). *Revista de Comunicación y Salud*, 10(2), 549-568. [https://doi.org/10.35669/rcys.2020.10\(2\).549-568](https://doi.org/10.35669/rcys.2020.10(2).549-568)
- Thelwall, M. & Levitt, J. M. (2020). Retweeting Covid-19 Disability Issues: Risks, Support and Outrage. *El Profesional de la Información*, 29(2). <https://doi.org/10.3145/epi.2020.mar.16>
- Thelwall, M. & Thelwall, S. (2020). Covid-19 Tweeting in English: Gender Differences. *El Profesional de la Información*, 29(3). <https://doi.org/10.3145/epi.2020.may.01>
- Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., ... Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *PLOS ONE*, 10(6), e0129179. <https://doi.org/10.1371/journal.pone.0129179>
- UNDP. (2020). Understanding what data tell us about COVID-19's socio-economic impact. Retrieved on May 21, 2021 from <https://www.undp.org/blogs/understanding-what-data-tell-us-about-covid-19s-socio-economic-impact>

- van Dijck, J. (2016). *La cultura de la conectividad: Una historia crítica de las redes sociales* (The Culture of Connectivity: A critical History of Social networks). Buenos Aires, Argentina: Siglo XXI.
- Vega, F. (2020, March 24). Los medios tradicionales recuperan poder y credibilidad con la pandemia provocada por el Coronavirus (Traditional Media Recover Power and Credibility with the Pandemic Caused by the Coronavirus). *Comscore*. Retrieved from <https://www.comscore.com/lat/Prensa-y-Eventos/Blog/Los-medios-tradicionales-recuperan-poder-y-credibilidad-con-la-pandemia-provocada-por-el-Coronavirus>
- Wikström, P. (2014). #srynotfunny: Communicative Functions of Hashtags on Twitter. *SKY Journal of Linguistics*, 27, 127-152. Retrieved from <http://www.linguistics.fi/julkaisut/SKY2014/Wikstrom.pdf>

AGRADECIMIENTOS

La Dra. Sued agradece al Programa de Estancias Postdoctorales por México 2020-2021 y al Sistema Nacional de Investigadores, Conacyt México, por el apoyo a la realización de este artículo.

SOBRE LOS AUTORES

DR. MANUEL CEBRAL LOUREDA, Doctor en Filosofía por la Universidad de Santiago de Compostela (España) y Maestro en Minería de datos por la UNED (España). Desde 2019 ejerce como profesor de planta en el Tecnológico de Monterrey (México). Sus intereses giran en torno a las Humanidades digitales y la cultura digital combinando el acercamiento humanístico con métodos basados en datos a través de herramientas de computación como la programación en R. Actualmente es miembro del Sistema Nacional de Investigadores (México) a nivel Candidato.

 <https://orcid.org/0000-0001-6359-2427>

DRA. GABRIELA ELISA SUED PALMEIRO, Doctora en Estudios Humanísticos por el Tecnológico de Monterrey, México. Es Investigadora postdoctoral Conacyt en el Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México. Practica la docencia de grado y postgrado en el Tecnológico de Monterrey. Sus líneas de investigación se centran en la Cultura Digital, los estudios de plataformas sociales, y los métodos centrados en datos.

 <https://orcid.org/0000-0002-4516-678X>