

APROXIMACIONES METODOLÓGICAS PARA LA INVESTIGACIÓN DIGITAL EN TWITTER

Luis Phillips

Boliviano. Magister en Ciencias de la Geoinformación y Observación de la Tierra. Sociólogo. Investigador social y docente de la Universidad Católica Boliviana “San Pablo”.

Código ORCID: 0000-0002-1656-4949

luis.phillips@ucb.edu.bo

Alex Ojeda

Boliviano. Magister en Investigación en Ciencias Sociales. Sociólogo. Investigador social y docente de la Universidad Católica Boliviana “San Pablo”.

Código ORCID: 0000-0001-6715-6825

aojeda@ucb.edu.bo

Andrea Alemán

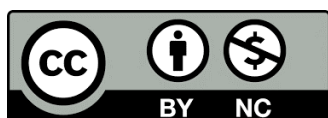
Boliviana. Magister en Desarrollo Global y Planificación. Comunicadora Social y Socióloga. Docente de la Universidad Católica Boliviana “San Pablo”.

Código ORCID: 0000-0001-5914-0364

aaleman@ucb.edu.bo

Los autores declaran no tener conflicto de interés alguno con la revista Punto Cero

Phillips, Luis. Ojeda, Alex. Alemán, Andrea. (2021). Aproximaciones metodológicas para la investigación digital en Twitter. Punto Cero, año 26 n°43, Diciembre de 2021. Pp 40-54. Universidad Católica Boliviana “San Pablo” Sede Cochabamba.



Resumen:

Como parte de la investigación “Sentidos de democracia generados en las comunidades digitales conformadas en Twitter, a partir de la crisis del 20 de octubre de 2019 en Bolivia”, el artículo tiene por objetivo introducir elementos metodológicos esenciales para la consulta de datos desde la Application Programming Interface (API) de Twitter, por medio del IDE de RStudio y las librerías “academictwitteR” y “tidyverse” aplicadas en el estudio. Se hace un recorrido por el modelo de datos que Twitter maneja para optimizar las consultas en su API v2 y el Academic Research Product Track, que permite a las investigaciones académicas acceder a la base histórica de tweets y realizar consultas. Además, se presentan tres vertientes de análisis y visualización de los datos obtenidos: el análisis estadístico descriptivo, el análisis de contenido y el análisis de redes. La convivencia de la sociedad entre el Internet y la esfera presencial resalta la necesidad de aplicar métodos sociales digitales para el análisis de datos. En consecuencia, este artículo se convierte en un espacio para la comprensión de un método aplicado a este tipo de investigaciones.

Palabras clave:

Investigación social, redes sociales digitales, medios sociales, Twitter

Methodological approaches for digital research on Twitter

Abstract:

As part of the research “Senses of democracy generated in the digital communities formed in Twitter, as of the crisis of October 20, 2019 in Bolivia”, the article aims to introduce essential methodological elements for consulting data from the API from Twitter through the RStudio IDE and the “academictwitteR” and “tidyverse” libraries applied in the study. A tour is made of the data model that Twitter uses to optimize queries in its API v2 and the Academic Research Product Track, which allows academic research to access the historical base of tweets and make queries. In addition, three aspects of analysis and visualization of the data obtained are presented: descriptive statistical analysis, content analysis and network analysis. The coexistence of society between the Internet and the face-to-face sphere highlights the need to apply digital social methods for data analysis, which is why this article becomes a space for understanding a method applied to this type of research.

Key words:

Social research, digital social networks, social media, Twitter

1. Introducción

Cada vez es más latente la vida digital de los bolivianos. En el ranking Alexa (2021), se hallan entre los sitios web más visitados las plataformas digitales como YouTube, Facebook y Live de Microsoft; portales de noticias como El Deber y Unitel; sitios oficiales del gobierno como la del Ministerio de Educación y diversos espacios de entretenimiento audiovisual. Estos hacen visible el abanico de espacios digitales concurridos en Bolivia para satisfacer necesidades de información, educación, entretenimiento y política, entre otras. Si bien esta vida digital fue consolidada por la pandemia, los sucesos políticos en nuestro país a fines de 2019 sellaron un hito a través del Internet que marcó una notoria huella digital y presencial que aún repercute.

Además de Facebook y WhatsApp, fue Twitter la red social digital que visibilizó los polos ideológicos en debate a una escala internacional durante los conflictos de octubre y noviembre, durante ese lapso de postelecciones presidenciales. Ello se debió, en gran medida, al uso de este espacio por el expresidente Evo Morales Ayma, quien, por medio de Twitter, enviaba mensajes tanto a los bolivianos cuanto a la comunidad internacional sobre los sucesos nacionales.

Asimismo, como un claro ejemplo de ese debate en Twitter, se tiene la pelea de los hashtags #GolpeDeEstado y #FraudeEnBolivia, a causa de que el 22 de octubre de 2019 Morales mencionó por primera vez un golpe de Estado que veía latente (Phillips y Alemán, 2019). Este hecho resalta la cualidad de Twitter de una red social digital que se relaciona a ejercicios ciudadanos y políticos. En palabras de Congosto, Fernández y Moro: “Twitter es una fuente de información que permite

segmentar a los usuarios, descubrir cómo los ciudadanos participan en el debate político y cómo se agrupan por afinidad ideológica” (Congosto y otro, 2011, p. 12).

Como señalan los autores, esta red social digital viabiliza en su concisión de 280 caracteres (antes 140) la expansión del mensaje a través de la retransmisión de los followers sin ninguna limitación, convirtiéndose en “una de las mayores fuentes públicas de propagación de la información a tiempo real” (2011). Si bien los mensajes son breves, su expansión puede tener un alcance de millones de usuarios; esta situación lleva a tomar decisiones metodológicas para su estudio ligadas al uso de la tecnología.

El gran movimiento de mensajes y el alto contenido de datos y usuarios, en esta red social digital, exigen investigaciones que hagan uso de métodos de recolección, procesamiento y visualización de datos que apliquen instancias y programas informáticos como la API de Twitter, Postman, RStudio; librerías como `academictwitteR` o `tidyverse` y visualizadores como Gephi o VOSviewer, NodeXL, entre otros. Si bien estudios en Twitter a nivel internacional hacen uso de estos recursos para analizar la ciberpolítica a nivel latinoamericano y en Bolivia (Manfredi-Sánchez y otro, 2021; Villa-Cox y otro, 2021; Waisbord y Amado, 2017), a nivel nacional, son escasos los artículos y las investigaciones que trabajan con estos métodos y en esta red social digital.

En ese sentido, este artículo apunta a explicitar y presentar un método alternativo para encarar estudios sociales en Twitter, con el objetivo de introducir elementos metodológicos esenciales para la consulta de datos desde la API de Twitter. Esta metodología se aplicó en una investigación digital cuyo objetivo

fue establecer los sentidos de democracia generados en las comunidades digitales conformadas en Twitter a partir de la crisis del 20 de octubre de 2019 en Bolivia (septiembre 2019 a noviembre 2020), luego de las elecciones presidenciales.

Este artículo, fruto de los avances en investigación, presenta indicios de los resultados obtenidos resaltando, sobre todo, las posibilidades de análisis digital que se pueden aplicar a estudios similares. En ese sentido, desarrolla en detalle el diseño de investigación social digital y las alternativas de análisis de datos, enfatizando y aclarando los elementos para este proceso y el lenguaje técnico que debe aplicarse. Por último, el artículo otorga, además, pautas a seguir con el fin de realizar investigaciones digitales académicas específicamente en Twitter.

2. El diseño de investigación digital en Twitter

En el acápite, se exponen las principales herramientas utilizadas para la recolección y procesamiento de los datos obtenidos de las redes sociales digitales, en este caso en particular, provenientes de Twitter. Asimismo, se presentan los elementos fundamentales que hacen al modelo de datos de Twitter, los entornos y niveles de consulta, los objetos que pueden ser consultados, los productos que tiene Twitter para los desarrolladores involucrados en investigación académica y la forma de acceder a ellos. Con ello, se ingresa a la descripción de un procedimiento para realizar consultas por medio de lenguaje de programación, siguiendo un flujo de trabajo que logra concluir con una base de datos de calidad para su procesamiento.

2.1. Twitter API v2

Para realizar consultas a la base de datos de Twitter, esta plataforma ha desarrollado una interfaz de programación de aplicación API v2, a través de la cual se puede realizar peticiones de información de acuerdo a los intereses del usuario, utilizando un nivel de lenguaje sencillo y de alto nivel. Una API es un conjunto de protocolos y definiciones que hace posible la conexión remota entre distintas instancias de software (IBM, 2021); es un intermediario que facilita la realización de consultas e ingreso de la información desde distintos puntos de acceso, administrando la interacción con otras instancias y posibilitando el desarrollo de aplicaciones dentro de una red de comunicaciones.

La manera en que Twitter comparte su información es a través de pedidos realizados desde la API de Twitter (o interfaces enlazadas a ella). Las consultas, solicitudes o pedidos de información (request) parten de una variedad de objetos disponibles. Estos objetos contienen campos de consulta para los distintos elementos que los conforman; entonces, las peticiones son construidas con base al modelo de datos de Twitter y permiten, mediante la asignación de parámetros de búsqueda, la segmentación del tipo de información disponible para consulta. Por último, las solicitudes son ejecutadas en función al tipo de información que se desea coleccionar y, dependiendo de ello, se realiza la consulta a un endpoint concreto. Cuando la API de Twitter interactúa con otro sistema, el punto de contacto entre estos es considerado un endpoint. Por lo tanto, los endpoints son identificados como los extremos de un canal de comunicación. A manera de ejemplo, cuando se realiza un pedido de información a través de la API de Twitter, en el momento en que este se contacta

con la base de datos de Twitter, dicho punto de contacto se considera un endpoint, y, dependiendo del tipo de consulta, devolverá como respuesta un tipo de recurso concreto. Las solicitudes pueden ser dirigidas a los siguientes endpoints (Twitter, 2020a): a) el endpoint de búsqueda reciente (recent search), para consultar información sobre los tweets públicos más recientes, hasta siete días previos a la solicitud, b) el endpoint de muestra de flujo (sampled stream), que devuelve el 1% de los Tweets publicados en tiempo real a la solicitud, c) el endpoint de flujo filtrado (filtered stream), el cual permite, aplicando operadores, seguir en tiempo real cómo se desarrolla algún evento, d) el endpoint de búsqueda de Tweets (Tweet lookup), el cual, a partir de IDs de Tweets, devuelve información sobre los mismos, y, por último, e) el endpoint de búsqueda de usuario (user lookup), el cual, a partir de IDs de usuario, devuelve información sobre ellos (Twitter, 2020b).

La API de Twitter maneja un lenguaje de alto nivel para los endpoints. Ello permite acceder a conjuntos de información más que a elementos separados. Primero, se realiza la consulta a través de los recursos de alto nivel y, luego, a partir de campos y extensiones, se extrae el resto de la información. Cada endpoint maneja un tipo de recurso de alto nivel (top-level resources) en el momento de procesar el pedido y responder a la API. Entre los principales, tenemos a los usuarios o los mismos tweets, que serán los elementos iniciales para consultar, dependiendo del objetivo de la búsqueda. Luego, el pedido, dependiendo de los parámetros de búsqueda, entregará información adicional hasta llegar a cierto nivel de agregación. Por ejemplo, cuando se quieren extraer los datos de métricas (likes, retweets, quotes, mentions, etc.), el pedido se hace una vez por todas las métricas y no así por un elemento en específico. Esto permite más información y menos escritura para acceder a estos datos.

Dentro de su modelo de datos, Twitter dispone de seis tipos de objetos que pueden ser consultados para el análisis (Twitter, 2020a). El primero, y más importante, es el tipo de objeto Tweet (tweet object): Es el elemento central de las interacciones dentro de la plataforma y contiene el conjunto de elementos presentes en una publicación. El segundo tipo de objeto es el de usuario (user object), el cual contiene todos los metadatos de una cuenta de Twitter. El tercer tipo de objeto se denomina space (space object) que permite la interacción y expresión de contenido por medio de conversaciones de audio en vivo, aunque sus características difieren un tanto de los tweets ya que están disponibles de manera más efímera y mientras duren las transmisiones. El cuarto tipo de objeto es el de contenido multimedia (media object), relacionado con cualquier imagen, GIF o video añadido a un tweet; este, a diferencia de los anteriores no es un objeto primario, pero puede ser consultado como campo de expansión de algunos de los objetos previos. El quinto tipo de objeto, al igual que el anterior, no es un objeto primario, pero puede ser consultado como campo de expansión; se trata del tipo de objeto encuesta breve (poll object), puede ser integrado a un tweet y su consulta posibilita ver la encuesta breve, la duración de la consulta y los resultados obtenidos. Por último, el tipo de objeto de lugar (place object) permite la clasificación de un tweet a partir del lugar de publicación del tweet o de los lugares etiquetados dentro de uno. Esta información, dependiendo de las expansiones consultadas, cuenta con varios niveles de agregación, desde países hasta un punto geolocalizado en coordenadas. El objeto de lugar no es un objeto de tipo primario y debe de ser consultado a través de campos de expansión.

2.1.1. Tipo de objeto Tweet (Tweet Object)

Un Tweet es el bloque básico del contenido publicado en Twitter, puede contener 280 caracteres y ser publicado de forma pública o privada, dependiendo de la cuenta (Twitter, 2020b). Es un conjunto de elementos, los cuales, dentro de un límite máximo de caracteres, construyen un mensaje. Este objeto está compuesto por varios campos (fields) representados en dos niveles: el nivel parental o nivel raíz (root-level) y el nivel de expansiones (child-level), el cual añade información al nivel parental. Al realizar un pedido de información, el nivel parental se encuentra por defecto incluido en la respuesta, mientras que, para consultar el nivel de expansiones, se debe incorporar parámetros de campo para realizar el pedido de dicha información.

Asimismo, el nivel parental de un tweet está compuesto por tres campos: el identificador (id), el texto (text) y la fecha de creación (created_at). Las principales expansiones de campo hacen viable la obtención de información sobre entidades asociadas (hashtags, cashtags, menciones, url, etc.), identificadores asociados (usuarios, conversación, localización), métricas (likes, replies, retweets, quotes, clicks, etc.), al igual que los datos de referenciación (en el caso de replies y quotes) y otra información de contexto.

2.1.2. Tipo de objeto Usuario (User Object)

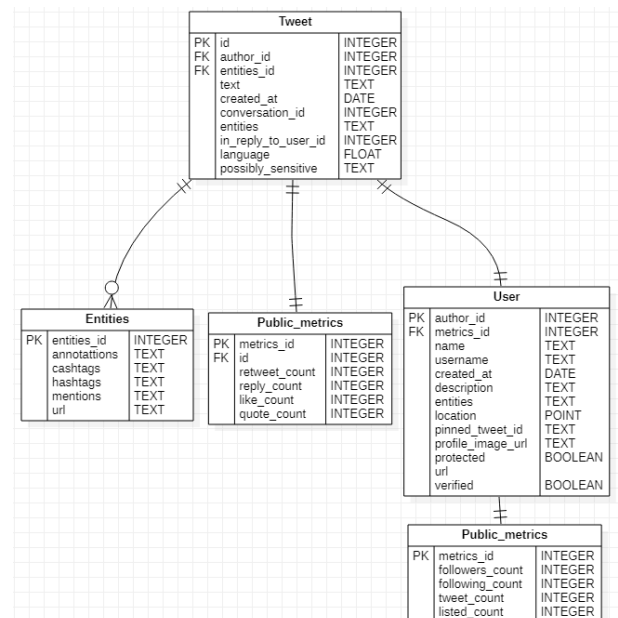
Este objeto contiene los metadatos asociados a un usuario. Consta del conjunto de información alojada en el sistema que facilita la identificación de los datos de un perfil dentro de Twitter. Esta información parte, en el nivel parental, del identificador (id) asociado al usuario, siendo posibles de consultar el resto de los campos dentro del nivel de expansiones. Este objeto puede ser consultado en dos niveles: de manera parental, pidiendo datos asociados a un identificador, o como

campo de expansión de una consulta del objeto tipo tweet, añadiendo los datos de usuario al tweet en cuestión.

Entre los principales campos de usuario expandibles, tenemos el nombre de usuario definido para el perfil (name), el alias o nombre actual en pantalla (username), la fecha de creación de dicho usuario (created_at), la descripción o bio del perfil (description), las entidades asociadas al usuario provistas en la descripción (entities), la localización (location), la imagen de perfil (profile_image_url) y las métricas de usuario (followers, following, tweet_count, listed_count), entre otros.

A continuación, presentamos un diagrama para visualizar con mayor claridad las relaciones entre los elementos, según los tipos de objeto.

Figura 1. Diagrama de entidades y relaciones (ERD)



Fuente: Elaboración propia

El diagrama muestra una visión simplificada de las relaciones entre elementos de un tweet. Esta visualización

permite aproximarnos a la identificación de los elementos principales a solicitar en las búsquedas, además de reconocer los outputs y, mediante el tipo de dato, planificar el procesamiento y análisis de la información.

2.2. Academic Research Product Track

Twitter ha lanzado en agosto de 2020 la versión 2.0 de su API. Junto con esta, anunció la implementación de un nuevo nivel de acceso para desarrolladores en proyectos de investigación académica (Academic Research Project). Este producto (Twitter, 2020c), presentado en enero de 2021, permite a los investigadores acceder a búsquedas en endpoints como el archivo completo de Twitter (full-archive search) o el archivo completo de conteo de tweets (full-archive Tweet counts), junto a funciones avanzadas de búsqueda. Se suman a ello el incremento de los parámetros de búsqueda y el número de consultas por mes, frente a otros productos ofrecidos por la empresa de manera comercial, ya que, es importante mencionar, una vez aceptada la postulación a este nivel de acceso, el mismo no conlleva un costo o importe debido a la finalidad académica que persigue y el uso no comercial de los resultados de investigación.

Los principales requisitos solicitados para acceder a este producto son: la pertenencia a una institución académica o universidad (acreditando dicha información), la presentación de un perfil de investigación (objetivos, metodología, finalidad y uso de la información), la apertura de una cuenta como desarrollador en la plataforma y la creación de un proyecto de investigación. Esta solicitud es procesada manualmente por personal de Twitter Developers, quienes verifican que los datos sean fehacientes y cumplan los requerimientos para otorgar el acceso. Al final, envían un mensaje

confirmando la solicitud o pidiendo la clarificación de algunas de las exigencias, en caso necesario. Una vez obtenido el permiso, en la plataforma de Twitter Developers se encuentran las claves de acceso imprescindibles para introducir en el momento de efectuar una solicitud desde la API. Los pasos para realizar consultas serán descritos a continuación, desde varios entornos y lenguajes de programación.

2.3. Recolección de datos

En este apartado, se presenta cómo realizar la solicitud de información por medio de un entorno de programación lo suficientemente flexible para obtener los datos que requiere la investigación. Los componentes más importantes a tomar en cuenta son: el flujo de trabajo (donde se abordan las principales consideraciones cuando se elabora una solicitud), el lenguaje de programación, el entorno para llevar a cabo el pedido y la sintaxis necesaria para hacer la consulta.

El flujo de trabajo (workflow), recomendado para solicitar datos en cualquiera de los entornos, debe conseguir realizar, como mínimo, los siguientes pasos: 1) autenticación, introducir y almacenar las credenciales de autorización (solo mediante su validación permitirá a la API realizar consultas), 2) claridad en la introducción de términos de búsqueda para una consulta eficiente y 3) correcto almacenamiento de los datos consultados. A continuación, presentamos un procedimiento para solicitar datos de la API de Twitter en un entorno de programación óptimo para Ciencias Sociales (RStudio) y a través de una librería de funciones (academictwitteR) que facilitan la consulta.

R es una plataforma para el lenguaje de programación orientado a objetos S/

S+, de código abierto (open source). Es de amplio uso para aplicaciones estadísticas, pero tiene un aporte importante para las Ciencias Sociales, sobre todo porque facilita la exploración de datos y su representación (Clark, 2014). Uno de los entornos aconsejados para desplegar R es RStudio, el cual es un IDE (Integrated Development Environment) para R. Incluye una consola, un editor de sintaxis que soporta la ejecución directa de código, además de herramientas de manejo de espacio de trabajo y de visualización de información.

Para realizar la recolección de información en este entorno, es necesario instalar y habilitar librerías que contengan las funciones necesarias en pro de dicho fin. En este caso, se hace uso de la librería “academictwitter”, la cual, a decir de sus autores, es un “paquete R para acceder el Academic Research Product Track v2 endpoint” (Barrie y Ho, 2021). Siguiendo el flujo de trabajo, se presentan los pasos de pedido de información por medio de RStudio.

2.3.1 Autenticación

Una vez creado un proyecto para la investigación en RStudio, el primer paso es instalar la librería “academictwitter” y activarlo para utilizar sus funciones. Una vez ejecutado, debemos introducir el o los tokens de autorización que proporciona Twitter Developers. A continuación, presentamos la sintaxis para realizar este paso:

Figura 2. Autenticación

```
1. install.packages("academictwitter") # Instalar la librería
2.
3. library(academictwitter) # Activar la librería
4.
5. set_bearer() # Comando para introducir y almacenar el código de autorización
6.
7. # Se abre una ventana ".Renviron" en la cual se debe introducir lo siguiente:
8.
9. TWITTER_BEARER=INTRODUCIR_BEARER_TOKEN # Almacena el código de autorización
10.
11. # Cerrar y reiniciar RStudio
```

Fuente: Elaboración propia

Con ello, tendremos almacenados los permisos necesarios para ejecutar las funciones de la

librería. Este paso solo debe ser repetido en caso de renovar los tokens de acceso.

2.3.2 Recolección

La recolección de datos se realiza por medio de la función “get_all_tweets()”, la cual permite recuperar información desde una búsqueda al endpoint del archivo completo de tweets (full archive search).

A continuación, se muestra un ejemplo de la sintaxis para realizar una búsqueda de 100 términos (tweets) que contengan las palabras “Bolivia” y “Democracia”, dentro de un periodo concreto de tiempo.

Figura 3. Recolección a partir de Tweets

```
1. library(academictwitter) # Al iniciar RStudio es necesario activar la librería
2.
3. # Definimos el nombre del output y la función que ejecutaremos, ingresando
   # los argumentos (o parámetros) a partir del diccionario de datos.
4.
5. tweets <-
6.   get_all_tweets(
7.     query = c("Bolivia Democracia"),
8.     start_tweets = "2019-01-10T00:00:00Z",
9.     end_tweets = "2019-31-10T23:59:59Z",
10.    is_retweet = FALSE
11.    file = "sentidos_democracia",
12.    data_path = "data/",
13.    bind_tweets = FALSE,
14.    n = 100,
15.  )
```

Fuente: Elaboración propia

Para más información, se presenta una breve descripción de cada uno de los argumentos utilizados en la sintaxis. Se incluye el nombre del argumento, el tipo de elemento y una breve descripción de lo que hace.

Tabla 1. Descripción de argumentos utilizados en la sintaxis

Argumentos	Tipo	Descripción
query	string, character vector	Elemento de consulta, p. ej. "democracia", "#Bolivia".
start_tweets	string	Fecha y hora inicial de la búsqueda
end_tweets	string	Fecha y hora final de la búsqueda
is_retweet	logical	Define si incluye en la respuesta objetos que sean Retweets
file	string	Nombre del archivo a exportar
data_path	string	Ruta donde se guardará el archivo
bind_tweets	logical	Define si el resultado es transformado a dataframe
n	integer	Límite de respuestas consultadas

Fuente: Elaboración propia

La correcta definición de los argumentos de consulta permitirá realizar solicitudes claras y precisas, reduciendo el tiempo de preprocesamiento y limpieza de la base de datos, pasos previos al procesamiento. Ahora, presentamos una solicitud con inclusión de usuarios. En el ejemplo, solicitamos que nos devuelvan los tweets (hasta 100) que contengan las palabras “bolivia” o “democracia” publicados en las cuentas de Twitter “@evoespueblo” (cuenta de Evo Morales) y “@carlosdmesag” (cuenta de Carlos Mesa, postulante a las elecciones 2019), durante el mes de enero de 2021.

Figura 4. Recolección a partir de usuarios

```

1. library(academictwitter)
2.
3. tweets <-
4.   get_all_tweets(
5.     query = c("bolivia", "democracia"),
6.     users = c("carlosdmesag", "evoespueblo"),
7.     is_retweet = FALSE,
8.     start_tweets = "2021-01-01T00:00:00Z",
9.     end_tweets = "2021-02-01T00:00:00Z",
10.    file = "analisis_mensajes",
11.    data_path = "data/",
12.    bind_tweets = FALSE,
13.    n = 100,
14.  )

```

Fuente: Elaboración propia

En esta solicitud, se ha añadido el argumento “users” el cual permite incorporar nombres de cuentas al pedido. Una posible variación que se puede hacer es la búsqueda solo por usuario, dejando el argumento “query” fuera de la sintaxis. Cabe notar que la función “c()”, IGUAL posibilita la conjunción de varios elementos del mismo tipo en una lista provisional.

Para concluir, es importante, para este segundo paso del flujo de trabajo propuesto, construir los términos de búsqueda a partir de las necesidades de información de la investigación. A mayor claridad, mejores datos se obtendrán.

2.3.3 Almacenamiento

La importancia del almacenamiento radica en la utilidad a futuro de la base de datos obtenida. Las etapas de preprocesamiento

y procesamiento de datos necesitarán bases de datos que respeten el modelo de datos requeridos por las librerías de procesamiento para ejecutar sus funciones. Con el fin de lograr la traducción de los datos de origen a formatos aptos para las librerías de análisis, es necesario transformar los objetos. Para ello, la librería “academictwitterR” dispone de algunas funciones que facilitarán este paso. Presentamos, a continuación, la sintaxis para transformar los archivos originales en formato JSON a dataframe y tidy para su procesamiento en RStudio.

Figura 5. Almacenamiento de datos

```

1. library(academictwitter)
2.
3. # Conversión del archivo de Tweets a dataframe. Permite unir varios archivos
  almacenados en la carpeta "data/".
4.
5. tweets <- bind_tweets(data_path = "data/")
6.
7. # Conversión del archivo de Usuarios a dataframe. Permite unir varios archivos
  almacenados en la carpeta "data/".
8.
9. users <- bind_tweets(data_path = "data/", user = TRUE)
10.
11. # Conversión del archivo de Tweets a tidy. Permite unir varios archivos para
  procesarlos con la librería Tidyverse.
12.
13. tweets_tidy <- bind_tweets(data_path = "data/", output_format = "tidy")

```

Fuente: Elaboración propia

Esto ayuda a consolidar la base de datos en un único archivo, basado en un modelo de datos apto para aplicar funciones de análisis, mediante otras librerías, como tidyverse, específicamente construidas a tal fin. Con ello, se garantiza la integridad de la base de datos y posibilita el avance a la siguiente etapa de la investigación.

3. El análisis de datos de redes sociales digitales

En virtud de lo expuesto previamente, se ha obtenido el material para el análisis de datos. En este apartado, se mostrará tres posibilidades para el análisis y visualización de datos: En primer lugar, el análisis estadístico descriptivo, útil para presentar frecuencias y proporciones en las variables de tipo numérico y categórico; en segundo lugar, el análisis de contenido para los datos

textuales que permite identificar palabras clave, y finalmente, en tercer lugar, el análisis de redes que visualiza estructuras de interacción.

3.1 Análisis estadístico descriptivo

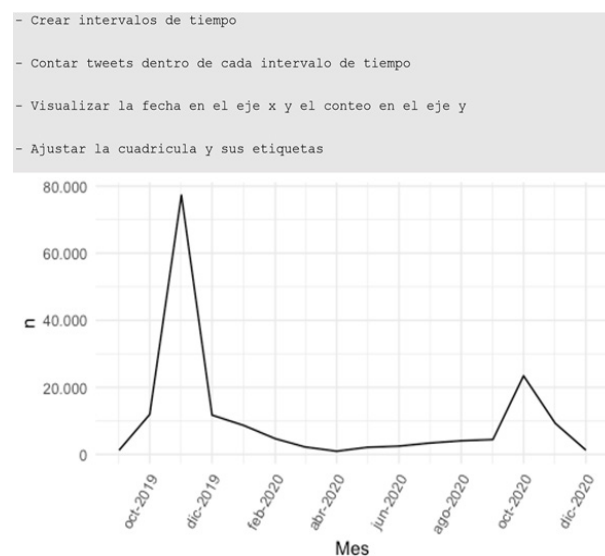
El análisis estadístico suele ser el primer paso exploratorio esencial para dar sentido a una gran cantidad de datos. La estadística en general posibilita “recolectar, organizar, resumir, presentar y analizar datos” (Spiegel y Stephens, 2009, p. 1); a su vez, la estadística descriptiva permite, principalmente, resumir la información, mostrando tendencias generales de los datos de redes sociales a través de técnicas sencillas de agregación y conteo.

3.1.1. Frecuencia de Tweets en el tiempo

Entre los datos por defecto que tiene un tweet y que también ofrece la API académica de Twitter, se encuentra la fecha de creación del tweet (`created_at`). Este dato es útil para facilitar la frecuencia de emisión de tweets en el tiempo, otorgando una idea del volumen de comunicación o hasta la posible viralidad de un tema de conversación y su asociación con determinados eventos contextuales. Facilita, además, identificar los picos de conversación.

Para visualizar la frecuencia de tweets, se puede aplicar un gráfico de línea, muy útil para visualizar tendencias en el tiempo.

Figura 6. Frecuencia de Tweets que contienen las palabras *democracia* y *Bolivia*



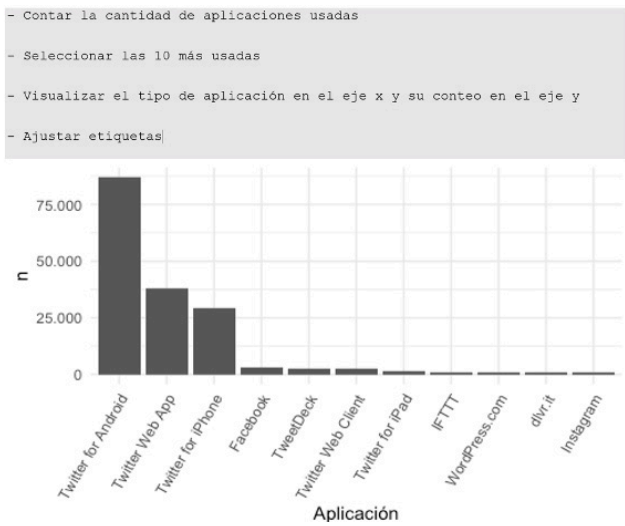
Fuente: Elaboración propia

Es fundamental definir con claridad el tamaño de los intervalos (año, trimestre, mes, semana, etc.), que dependerá de la ventana de análisis de más interés. En este caso, se optó por el mes. Se puede apreciar una mayor cantidad de tweets en los meses clave asociados con las coyunturas políticas y electorales bolivianas.

3.1.2 Porcentaje de fuentes de emisión de Tweets

La variable fuente (`source`) brinda la información de las aplicaciones usadas para emitir los tweets. La podemos visualizar con un diagrama de barras.

Figura 7. Tipo de aplicaciones usadas para emitir Tweets sobre democracia



Fuente: Elaboración propia

En este caso, se usaron 317 tipos de aplicaciones. Visualizarlas todas oscurecería el gráfico; por lo tanto, solo se eligieron las 10 más utilizadas. Entre ellas, se ve que hay una preminencia por las aplicaciones móviles frente a la aplicación web.

3.1.3 Frecuencias de interacciones

Para la interacción, se cuenta con las variables de “me gusta” (like), retweet, cita (quote) y respuesta (reply). Se puede visualizar su evolución en el tiempo con un gráfico de línea con líneas múltiples:

Figura 8. Tipo de interacción con los Tweets en el tiempo

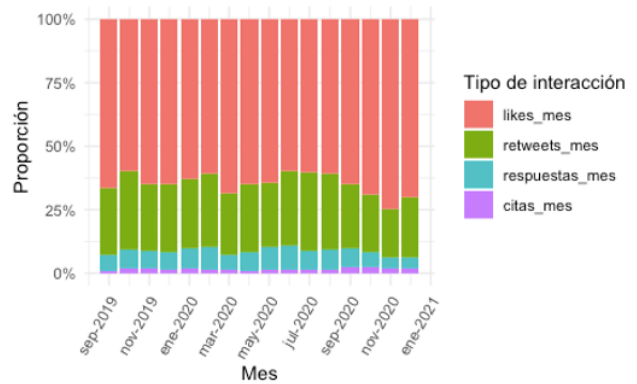


Fuente: Elaboración propia

En este caso, es muy llamativa la magnitud de las interacciones, principalmente los likes que llegan a más de dos millones en noviembre de 2019. Pero la calidad de las interacciones no es la misma; se ve en general que los likes y retweets, que requieren menos esfuerzo, son las principales interacciones frente a las respuestas a las conversaciones.

Adicionalmente, se puede contar con una visualización alternativa que haga énfasis en las proporciones de las interacciones entre sí en vez de solo conteos absolutos. Para ello, se utiliza un gráfico de barras apilado y normalizado.

Figura 9. Proporciones de las interacciones



Fuente: Elaboración propia

Se observan las mismas tendencias, pero ahora está más claro cuáles son las proporciones del total que ocupa cada tipo de interacción.

3.2. Análisis de contenido

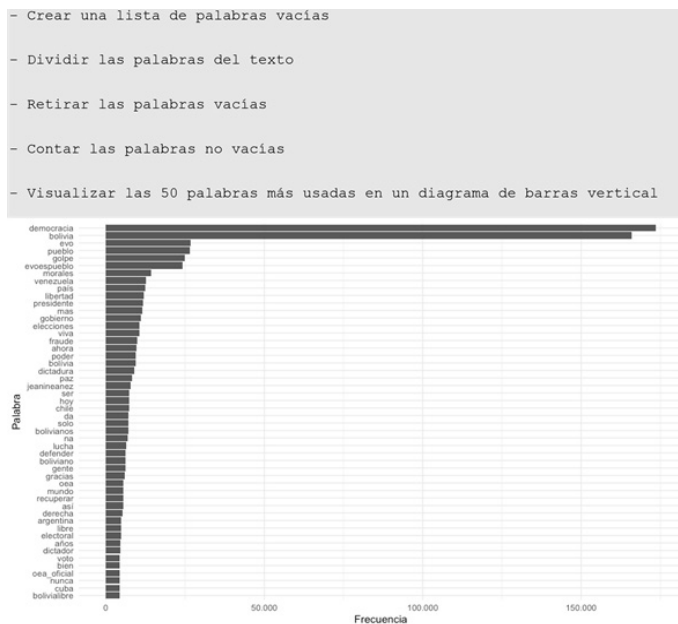
A diferencia de las variables categóricas y numéricas vistas en el anterior apartado, ahora se analizarán variables textuales. Para ello, se utilizan algunas técnicas del análisis de contenido. El propósito del análisis de contenido es “clasificar el material textual, reduciéndolo a conjuntos de datos más relevantes y manejables” (Weber, 1990). Esto es bastante útil, particularmente para el análisis de datos redes sociales, puesto que en muchas ocasiones la cantidad de texto emitido desborda las posibilidades de una lectura intensiva.

3.2.1. Frecuencia de Palabras

Entre los datos recolectados de la API académica de Twitter, se cuenta con la variable texto (text), que representa el contenido textual que se compartió en el tweet. Una primera técnica de análisis de contenido consiste en dividir el contenido textual en determinadas unidades textuales, como las palabras. Una vez realizada esa división, se puede contar automáticamente sus frecuencias. Esto lleva a una aproximación a las palabras clave más utilizadas y a los tópicos en tendencia.

Empezamos por las frecuencias generales en todo el período:

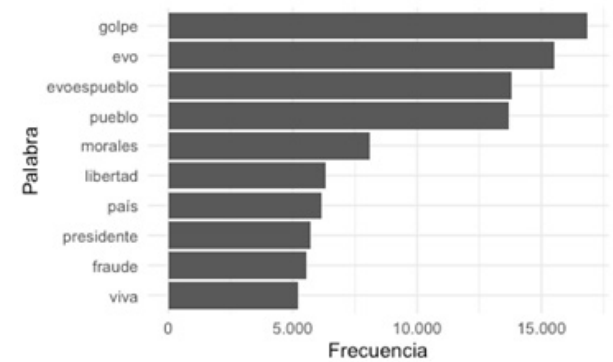
Figura 10. Las 50 palabras más utilizadas en los Tweets sobre democracia y Bolivia



Fuente: Elaboración propia

También, se puede realizar esta misma visualización por periodos mensuales. Para este caso, se comienza con el mes de más tweets e interacción (noviembre de 2019):

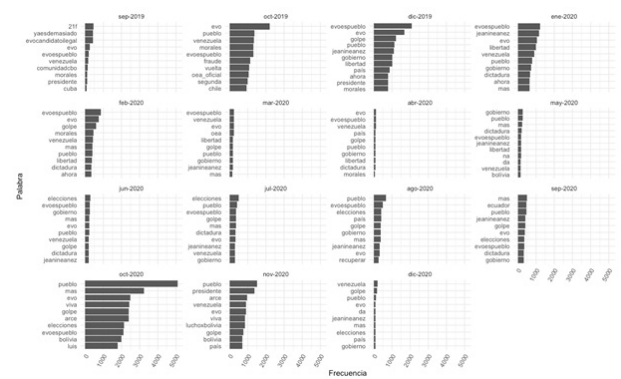
Figura 11. Las 10 palabras más usadas en los Tweets de 2019



Fuente: Elaboración propia

Asimismo, se puede visualizar las palabras más utilizadas cada mes durante el periodo de estudio (de septiembre a 2019 a diciembre de 2020, excluyendo noviembre de 2019):

Figura 12 Las palabras más usadas por mes



Fuente: Elaboración propia

Esto permite observar las constantes y los cambios en los tópicos de conversación durante el periodo de análisis.

3.3. Análisis de redes de conversación

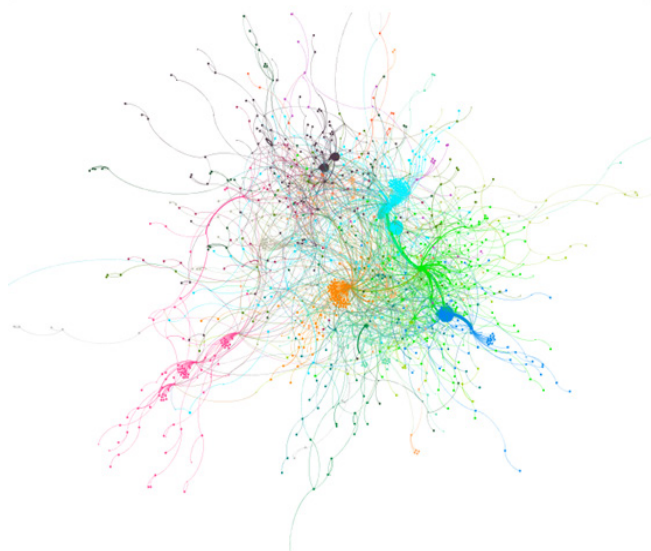
Un tercer tipo de análisis es el que ofrece el “análisis de redes sociales”.

Este es “un método estructural-relacional, formal y cuantitativo que se interesa

por los patrones de interacciones entre los actores” (Ojeda, 2015). Este método es útil para captar algunas propiedades formales de la interacción. Con los campos identificación de autor (author_id) y su respuesta (in_reply_to_user_id), se puede visualizar la forma de las conversaciones utilizando una red. Cada vértice será un usuario y cada arista una respuesta de un usuario a otro, con la posibilidad de que la arista tenga una magnitud en función a la cantidad de respuestas que se hayan realizado.

Figura 13. Red de conversación

- Crear tabla de vértices filtrando los usuarios únicos
- Crear la tabla de aristas creando dos variables, una de origen y otra de destino
- Calcular el peso de las aristas
- Filtrar a solo los usuarios que emitieron más de dos respuestas
- Filtrar al componente más interconectado
- Visualizar cada identificar de usuario como un vértice o nodo (punto)
- Visualizar cada respuesta de un usuario a otro mediante una arista (línea)
- Detectar subgrupos o comunidades internas
- Usar un algoritmo de disposición para desenredar la red (Force Atlas 2)



Fuente: Elaboración propia

En el gráfico, se pueden ver diferentes colores que representan a diferentes subgrupos o comunidades de conversación. Además,

existen ciertos vértices más grandes que otros en función a su cantidad de interacciones. Estas dos técnicas sirven para identificar tanto los grupos de conversación cuanto los líderes de opinión en cada subgrupo o, incluso, los vértices puente entre subgrupos.

4. 3 Discusión

Es evidente que la sociedad ha trasladado su hacer civil público o, al menos, ha llegado a la convivencia de este en espacios digitales. Las interacciones digitales como “nuevas formas de comunicación” se diferencian de las tradicionales por las particularidades de digitalización, reticularidad, hipertextualidad, multimedialidad e interactividad (Scolari, 2013), las cuales permiten procesos propios de interacción política en internet con una potente generación de datos.

Ello conlleva a la búsqueda de materiales y métodos para realizar investigaciones sociales digitales en pos de la recolección, procesamiento y visualización de datos. Estudios relacionados a comunicación y política, como la investigación que se halla detrás de este artículo metodológico, no pueden desligarse de esta realidad. El abanico de las técnicas tradicionales ya no es suficiente para la lectura de los procesos vivos en espacios digitales.

Es así como las redes sociales digitales producen una cantidad de datos sin precedentes. En el caso de Twitter, para el año 2020, se generaron, de manera global, en promedio, alrededor de 6.000 tweets por segundo (Sayce, 2019). Esto conlleva la acumulación de datos pasibles a ser analizados para entender las dinámicas que acontecen dentro de la plataforma y que, desde un panorama generalizador, presentan los acontecimientos y la discusión sobre temas que importan a los usuarios.

En este sentido, Twitter ha identificado a los investigadores académicos como uno de los grupos más grandes e importantes de desarrolladores que utilizan la API de Twitter (Tornes y Trujillo, 2021), realizando investigación en servicio de la comunidad a partir de los datos obtenidos en la plataforma (Twitter, 2021). La oportunidad de acceder al registro público de datos favorece las labores investigativas y, en el caso de las Ciencias Sociales, es una ventana de oportunidad para avanzar en la recolección de datos e investigación sobre temáticas relevantes y acontecimientos que impactan en las comunidades de estudio. Entonces, el Academic Research Product Track es un aporte fundamental para la investigación social digital, ya que posibilita fácilmente el acceso a los datos y a elaborar consultas específicas para profundizar en temas de interés para la investigación social.

Por otra parte, obtener información en grandes volúmenes y formatos digitales involucra el uso de herramientas digitales para su recolección y procesamiento, buscando optimizar los tiempos y recursos disponibles para la investigación. Estas herramientas hacen posible agilizar los procesos a través de paquetes, librerías y funciones desarrolladas en un lenguaje de alto nivel, diseñadas para automatizar, en gran medida, muchas de las tareas en varias etapas de la investigación. Esto es posible mediante entornos de programación y lenguajes con sintaxis sencillas e intuitivas, como R y RStudio, y librerías como “*academictwitteR*” o “*tidyverse*”, siendo que, una vez diseñadas las rutinas, pueden ser ejecutadas cuantas veces sean necesarias solo con la modificación de los argumentos, optimizando el tiempo dedicado a estas tareas.

Una vez obtenidos los datos, las técnicas de análisis y visualización descritas ayudan a dar sentido a ese gran volumen de información que por sí mismo no expresa con claridad su estructura y contenido. Si bien todavía

es imprescindible utilizar las técnicas tradicionales de lectura e interpretación contextual de las Ciencias Sociales, también se hace necesario aplicar técnicas digitales en el campo de la investigación de redes sociales digitales, precisamente por esa gran cantidad de volumen, velocidad y variedad de producción de la información. Las técnicas de análisis estadístico descriptivo y las de redes facultan la aproximación a la estructura de la comunicación digital; mientras que las técnicas de análisis de contenido posibilitan un acercamiento a los sentidos y tópicos de conversación. Los tres grupos de técnicas son complementarios. Estas técnicas, junto al diseño de levantamiento de datos de Twitter, brindan una comprensión amplia de los procesos políticos y sociales presentes en este espacio. A diferencia de otras redes sociales digitales, Twitter se abre en gran medida a otorgar datos a la academia para la realización de investigaciones. Ello se convierte en una veta para explotar, trazándose como una oportunidad para comprender el análisis de datos y aplicarlo en investigaciones sociales.

Bibliografía

Alexa. (2021). Alexa. Top Sites in Bolivia. <https://www.alexa.com/topsites/countries/BO>

Barrie, C. y Ho, J. (2021). *academictwitteR*: An R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software*, 6(62), 3272. <https://doi.org/10.21105/joss.03272>

Clark, M. (2014). R y Social Science. *Journal of Statistical Software*.

Congosto, M. L.; Fernández, M. y Moro, E. (2011). Twitter y política: Información, opinión y ¿predicción? En Cuadernos de comunicación. 4. Comunicación política 2.0 (pp. 11-16). Evoca. Comunicación e imagen.

IBM. (2021, abril 9). What is an Application Programming Interface (API). <https://www.ibm.com/cloud/learn/api>

Manfredi-Sánchez, J.L.; Amado-Suárez, A. y Waisbord, S. (2021). Presidential Twitter in the face of COVID-19: Between populism and pop politics. *Comunicar*, 29(66), 83-94. <https://doi.org/10.3916/C66-2021-07>

Ojeda, A. (2015). Introducción al Análisis de Redes Sociales. Mente Maestra.umss.edu.bo

Phillips, L. A. y Alemán, A. (2019). #GolpeDeEstado y #FraudeEnBolivia: Un análisis de las comunidades digitales en la crisis del 20/O. *Punto Cero*, 24(39), 51-61.

Sayce, D. (2019). The Number of tweets per day in 2020. dsayce. <https://www.dsayce.com/social-media/tweets-day/>

Scolari, C. (2013). Hipermediaciones elementos para una teoría de la comunicación digital interactiva. Editorial Gedisa. Barcelona

Spiegel, M. y Stephens, L. (2009). Estadística (4ta edición) McGraw Hill. México

Tornes, A. y Trujillo, L. (2021). Enabling the future of academic research with the Twitter API. https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api

Twitter (2020a). Tweet object. <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet>

Twitter (2020b). Twitter API Documentation | Docs | Twitter Developer Platform. <https://developer.twitter.com/en/docs/twitter-api>

Twitter (2020c). Twitter API for Academic Research | Products. <https://developer.twitter.com/en/products/twitter-api/academic-research>

Twitter (2021). Success stories | Twitter Developer Platform. <https://developer.twitter.com/en/community/success-stories>

Villa-Cox, R.; Zeng, H.; KhudaBukhsh, A. y Carley, K. (2021). Exploring Polarization of Users Behavior on Twitter During the 2019 South American Protests. *ArXiv*, 1. <http://arxiv.org/abs/2104.05611>

Waisbord, S. y Amado, A. (2017). Populist communication by digital means: Presidential Twitter in Latin America. *Information, Communication & Society*, 20(9), 1330-1346. <https://doi.org/10.1080/1369118X.2017.1328521>

FECHA DE RECEPCIÓN: 23/08/2021

FECHA DE APROBACIÓN: 15/11/2021