

Las redes sociales son servicios web que permiten al usuario construir, crear un perfil público o para cierto tipo de público o ciertas personas con quienes interactuará. Es así como se forma esta estructura social y el estudio de estos se retoma a los años 30's con la creación de sociogramas para luego crearse la sociometría, precursora del análisis de las redes sociales y también de la psicología social. Hoy por hoy, el análisis de las redes sociales es un estudio interdisciplinario donde convergen varios campos, como las ciencias sociales, la matemática, estadística, entre otros.[2]

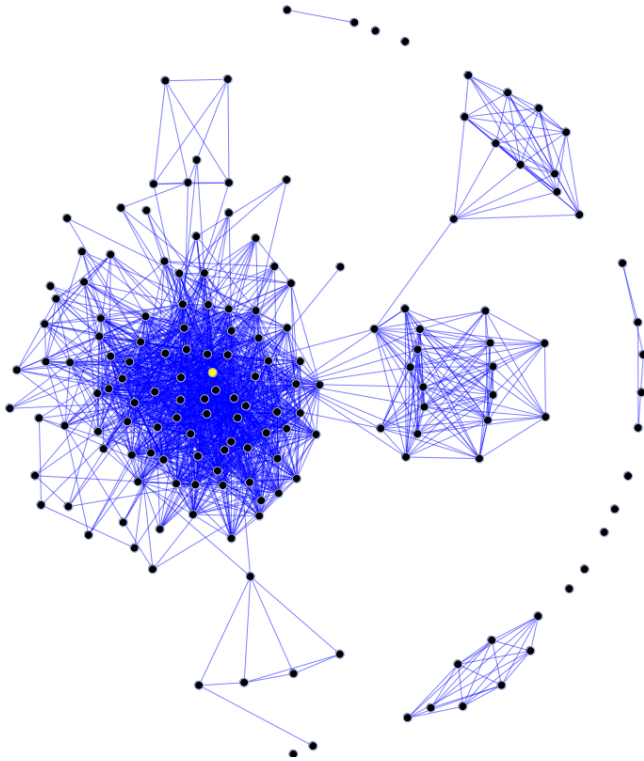


Figure 3: Estructura de una red social representada como grafo.[2]

3.1.1 Twitter. Twitter es una red social que permite la interacción de usuarios con sus seguidores de manera unidireccional. Cada uno de ellos tiene un perfil que puede o no ser público y de esta manera solo los seguidores de dicho usuario podrían ver su información y tweets o publicaciones. Dicha información puede ser compartida por otros usuarios a través del llamado retweet. Sin embargo, las publicaciones que se realizan en este medio están limitadas a un tamaño máximo de 280 caracteres, lo cual obliga a que los usuarios expongan sus ideas de la manera más sucinta posible, permitiendo que cada publicación tenga un significado en sí misma sin necesariamente estar ligada a otras para tener sentido. Es por lo práctico de la red, que es tan usada en el mundo.

3.2 Integración de datos basada en Api's

Las Api's son un conjunto de aplicaciones que permiten la construcción de una interfaz inteligente, configurando un medio por el cual dos sistemas se comunican. La principal ventaja de esta técnica es el hecho de que al ser servicios desarrollados con fines específicos, la velocidad es bastante óptima. Sin embargo, existe una desventaja muy considerable, las decisiones acerca de qué datos serán compartidos y el número de solicitudes al servicio que un usuario puede hacer, quedan estrictamente definidas por los sitios web, ocasionando que los análisis posibles de los datos sean limitados en cuanto a los aspectos que se podrían abordar o la cantidad de datos que se podría involucrar. [4]

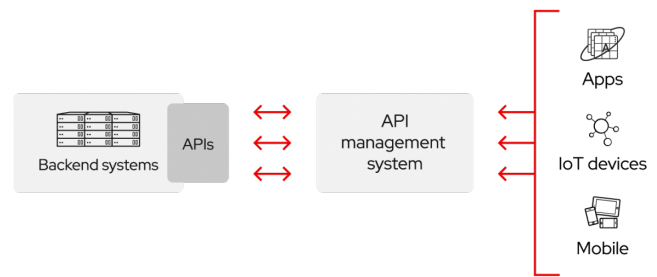


Figure 4: Estructura de una API.[5]

4 TRABAJOS RELACIONADOS

- Sentiment Analysis of Twitter Data[6] es un pequeño artículo que despertó la atención sobre el manejo de información de redes sociales en nosotros. Este pequeño artículo nos habla sobre análisis de sentimientos a través de tweets relacionados con donaciones, recaudación de fondos o organizaciones benéficas. Capturando el JSON, en este artículo se utiliza Python Twitter API para encontrar los ítems: donations, donor, charity, fundraising.
- Twitter Sentiment Analysis on Worldwide[3] COVID-19 Outbreaks, es un paper que a través de la biblioteca tweetpy de python y luego Textblob hace una búsqueda y representación gráfica de las coincidencias con las palabras: COVID-19 y coronavirus, donde los sentimientos más denotados fueron: deprimido y miedo, preocupación y aburrimiento. En menor medida aliviado o tranquilo.

Mediante el uso de la API de Twitter y la biblioteca tweepy de python. Se seleccionan dos palabras clave para buscar coronavirus y COVID-19. Los resultados mostraron que el costo neutral con respecto a las palabras clave de coronavirus y covid-19 para la polaridad fue significativamente alto, que es más del 50 por ciento, y la gran parte de los registros fueron objetivos, que fue de aproximadamente el 64 por ciento. De este estudio podemos decir que las reacciones de las personas varían día a día de publicar sus sentimientos en las redes sociales específicamente en Twitter.

5 PROPUESTA

A través de la integración de las API's, se puede observar por usuario cuánto interactúa determinada palabra o tema y así comparar quiénes tienen más influencia en las masas en dicho tema. De esta manera a través de la identificación de ítems frecuentes veremos qué usuario tiene más influencia sobre otros usuarios, ya sea para interactuar a favor o en contra, y con dicha información podremos observar la edad, fecha y lugar de los involucrados.

5.1 Método de Análisis

- Recopilación de información para el análisis de patrones en los influenciadores de redes sociales.
- Análisis de algoritmos propuestos
- Formulación del problema resolver.

- Elección de un patrón a analizar en Twitter con el fin de demostrar el impacto interactividad de ciertos usuarios sobre otros.

5.2 Algoritmo

Para diseñar nuestro algoritmo, hemos tomado en consideración algunos parámetros que nos devuelve la API al hacer la consulta, para este caso hemos considerado el número de listas públicas de un usuario, que básicamente es para ver que tan en común es un usuario dentro de un grupo, el número de retweets, el número de me gusta hacia un tweet y el número de seguidores de un usuario determinado.

Entonces tomando estas medidas, hemos hecho un cálculo para obtener la popularidad, dándole una ponderación a cada parámetro que se ha obtenido y de esa forma obtenemos una popularidad para un cierto tweet de una cierta persona o sitio, además esto lo hemos normalizado de 0 a 100 para que sea más fácil el manejo de los datos. Las medidas de ponderación que le hemos dado se muestran:

- listas públicas: 7
- interacción con un tweet: 5
- número de seguidores: 1

La fórmula para obtener la popularidad como tal, la mostramos a continuación:

$$x = \text{pond_listas} * \text{usuario}(\text{numero_de_listas})$$

$$y = (\text{pond_tweet_interacion} * \text{tweet_interaction})$$

$$z = \text{pond_seguidores} * \text{usuario}(\text{num_seguidores}) + 1$$

$$\text{popularidad} = x + y + z$$

Cabe resaltar que tweet_interaction se calcula las veces que un tweet ha sido retuiteado + las veces que recibió un me gusta

Entonces para realizar todo el proceso que se ha mencionado hasta el momento, se realizó código en el lenguaje de programación Python, con la ayuda de la librería Tweepy, y se pasa a detallar el procedimiento en código:

Listing 1: Importación de librerías y claves de acceso

```
1 import tweepy
2 import json
3 import pandas as pd
4
5 CONSUMER_KEY = 'twZ0LLfkkzHw2LeT349H6WjY6'
6 CONSUMER_SECRET = '
    SmVMH1KksU1VdGDHdMzqr0z0an5eIPq9rpIMKiI7P2dTSHd8tY
    '
7 ACCESS_TOKEN = '2707859331-
    x1cLXarR4JXFTh1j0G1TJZ1vcGmuYqDrY0I8UfK'
8 ACCESS_TOKEN_SECRET = '
    T7xT9QCmxPxiIPpfXGEnD07S3pDHut0809J5CydaKAXpp'
```

Una vez importadas las claves de acceso, que lo brinda Twitter, con una solicitud en su sitio web, se puede empezar a trabajar:

Listing 2: Aceptar o no ciertos tweets

```
1
2 def accept_tweet(tweet):
3     THRESHOLD = 3
```

```
4     interactions = tweet['retweet_count'] + tweet['
        favorite_count']
5     if interactions >= THRESHOLD and 'RT ' not in tweet
        ['full_text']:
6         return True
7     return False
```

En el código anterior, hacemos un filtro para no aceptar a cualquier tweet que devuelva la API, este filtro lo que realiza es si las interacciones han sido mayor a 3, esto se hace con el fin de ver quien es influyente en una noticia en particular, y también verificamos que no haya tenido un retweet, esto porque ya calculamos el número de retweets anteriormente y no queremos tener el mismo mensaje pero de un retweet.

Listing 3: Calcular popularidad

```
1
2 def calculate_popularity(tweet):
3     pond_public_list = 7
4     pond_tweet_interaction = 5
5     pond_follower_number = 1
6     user = tweet['user']
7     tweet_interactions = tweet['retweet_count'] + tweet
        ['favorite_count']
8     return (pond_public_list * user['listed_count']) +
        (pond_tweet_interaction * tweet_interactions)
        + (pond_follower_number * user['
            followers_count']) + 1
9
10 def normalize_popularity(popularity, min_popularity,
        max_popularity):
11     return 100 * (popularity - min_popularity) / (
        max_popularity - min_popularity)
```

Lo que realizamos anteriormente es lo que se explicó en la parte superior, el cálculo de la popularidad y la normalización para el cálculo de la popularidad.

Listing 4: Log

```
1
2 auth = tweepy.OAuthHandler(CONSUMER_KEY,
        CONSUMER_SECRET)
3 auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
4
5 api = tweepy.API(auth, wait_on_rate_limit=True,
        wait_on_rate_notify=True)
6 min_popularity = 0
7 max_popularity = 0
8 tweets = []
9 for tweet_parsed in tweepy.Cursor(api.search, q="#
        ParoDeTransportistas", tweet_mode="extended").items
        (7000):
10     tweet_str = json.dumps(tweet_parsed._json, indent=2)
11     tweet = tweet_parsed._json
12     if accept_tweet(tweet):
13         tweet_data = {
14             'user_id': tweet['user']['id_str'],
15             'user_name': tweet['user']['name'],
16             'popularity': calculate_popularity(tweet),
17             'user_location': tweet['user']['location'],
18             'tweet_msg': tweet['full_text']
19         }
20     tweets.append(tweet_data)
21     if max_popularity == 0:
22         min_popularity = tweet_data['popularity']
23         max_popularity = tweet_data['popularity']
24         continue
25     if tweet_data['popularity'] < min_popularity:
26         min_popularity = tweet_data['popularity']
27     elif tweet_data['popularity'] > max_popularity:
```

```
28         max_popularity = tweet_data['popularity']
29
30 tweets_analyzed = []
31 for tweet in tweets:
32     tweet['popularity'] = normalize_popularity(tweet['
33         popularity'], min_popularity, max_popularity)
34     tweets_analyzed.append(tweet)
35
36 tweet_str = json.dumps(tweets_analyzed, indent=2)
```

Lo que se realiza en esta parte es la extracción de los tweets en sí, y se pasa a tratar esos datos con las funciones que ya han sido mencionadas, para este caso el tema tratado es el paro de transportistas, y se evalúa la popularidad de un tweet, determinado y eso se pasa a un formato json, donde sus campos son el id del usuario, el nombre del usuario, la popularidad normalizada que calculamos nosotros, y el mensaje como tal, eso se hace para poder hacer un análisis gráfico.

6 EVALUACIÓN

Para realizar la evaluación generamos gráficos, en los cuales se puede apreciar la influencia de un determinado usuario para un determinado tema, esto dependiendo de su popularidad, por ello es muy importante resaltar eso, ya que la popularidad determina que tanto ha influido una persona o usuario en un tema dado.

Para hacer los gráficos se trabajo sobre un archivo csv, que lo que hace es mostrar la información de json pero en un formato de columnas para que los datos se traten de mejor manera y sea más fácil hacer los gráficos.



Figure 5: Datos en CSV.

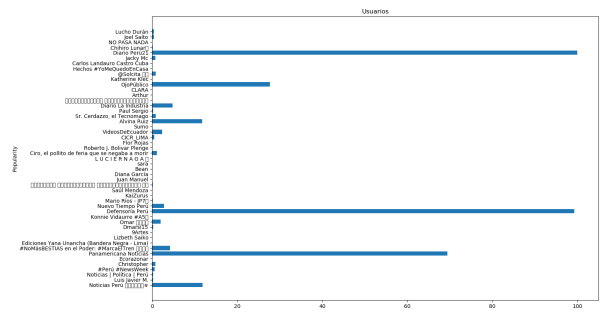


Figure 6: Influencia de los usuario hacia un tema.

Como se puede apreciar, el gráfico nos devuelve que usuarios han sido más influyentes en el tema del paro de transportistas, al ser demasiados usuarios su nombre no se puede apreciar de manera correcta, pero al hacer un acercamiento, tenemos para ello:

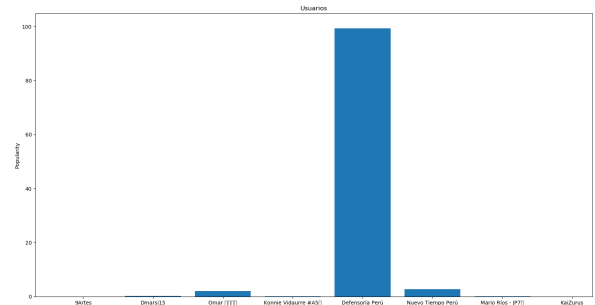


Figure 7: Usuario Defensoría Perú.

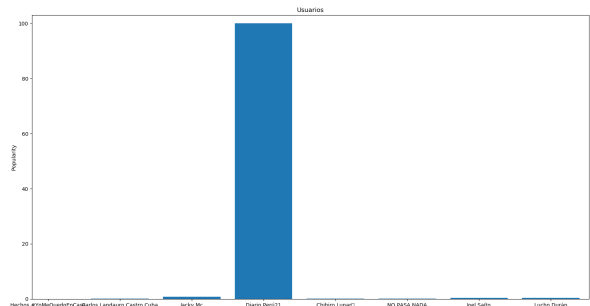


Figure 8: Usuario DiarioPeru21.

Una vez con los datos en ese formato, se procedió a hacer los gráficos correspondientes, que para este caso fue un gráfico de barras, donde se muestra la popularidad respecto a un usuario.

También vemos algunos otros usuarios que influyen pero no de la manera en que lo hacen los anteriores:

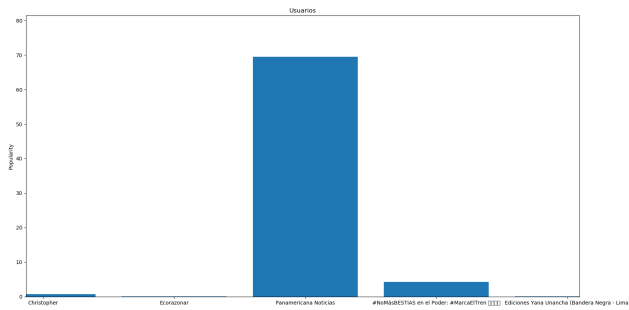


Figure 9: Usuario PanamericanaNoticias.

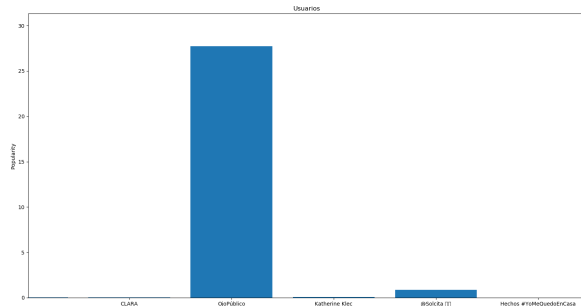


Figure 10: Usuario OjoPúblico.

Y como se ha podido apreciar se ve los que menos influyen, eso quiere decir que solo un grupo cercano de personas que lo sigue interactúa con sus tweets, pero no como lo hacen los anteriores usuarios.

7 CONCLUSIONES

- Nos dimos cuenta que la red social Twitter es un medio digital de comunicación muy usado por las personas, en donde la gente puede dar a conocer su opinión, y esta opinión puede llegar a ser muy influyente en otras personas.
- Nos dimos cuenta que si el sitio que publica o realiza tweets es un sitio de noticias, como diarios, programas televisivos, esto tiene más influencia en la gente, ya que puede considerar una fuente confiable para las noticias.
- Se pudo apreciar que aunque una persona tenga influencia, esta no va a ser tanta como la que tienen sitios verificados, esto pudimos apreciar también, los sitios verificados, o sea los que cuentan con una aspa a lado de su nombre, tienen mayor influencia en los usuarios.
- Twitter brinda para los desarrolladores una API, en la cual los desarrolladores pueden hacer consultas a un usuario, hacer filtros, entre otras características, la desventaja es que no se puede llegar a obtener toda la cantidad de tweets que ha ocurrido en el tiempo y mucho menos si son temas no tan influyentes.

- La red social bien usada, da la oportunidad de comunicar instantáneamente y masivamente temas de interés colectivo importantes y urgentes, dado su alcance potencial.

REFERENCES

- [1] es.qaz.wiki. 2020. Influencia social. Retrieved – from https://es.qaz.wiki/wiki/Social_influence
- [2] es.wikipedia.org. 2020. Redes Sociales. Retrieved – from https://es.wikipedia.org/wiki/Red_social
- [3] Pshko R. Mohammed Amin Kamaran H. Manguri and Rebaz N.Ramadhan. 2020. Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks. Video. ACM Press, Union, 4. <https://doi.org/99.9999/woot07-S422>
- [4] maplink. 2020. Integration API. Retrieved – from <https://maplink.global/es/blog/que-es-interfaz-programacion-aplicaciones-api/>
- [5] redhat. 2020. API. Retrieved – from https://www.redhat.com/cms/managed-files/styles/wysiwyg_full_width/s3/API-page-graphic.png?itok=5zMemp9
- [6] Amrita Shelar and Ching yu Huang. 2018. Sentiment Analysis of Twitter Data. Video. In *ACM SIGGRAPH 2003 Video Review on Animation theater Program: Part I - Vol. 145 (July 27–27, 2003)*. chool of Computer Science, Kean University, ACM Press, Union, 4. <https://doi.org/99.9999/woot07-S422>