

Basta analizar los retuiteos dentro de la hora posterior a su publicación, para saber si un tuit será trending topic



El grupo de Análisis y Procesamiento de Grandes Redes Sociales y Semánticas de la FaMAF-UNC analizó 9,5 millones de tuits. Centrándose en las conexiones sociales entre casi cuatro mil usuarias y usuarios, determinó que la primera hora de actividad de un posteo permite saber, con un 70% de efectividad, si será popular. Esta detección temprana posibilita, por ejemplo, crear contranarrativas que frenen el alcance de las fake news. [05.05.2021]





Por **Eliana Piemonte**

Redacción **UNCiencia**

Prosecretaría de Comunicación Institucional – UNC

eliana.piemonte@unc.edu.ar

Twitter tiene 353 millones de usuarias y usuarios activos en el mundo. Sus publicaciones tienen formato de titular y se leen rápido. Cuando a alguien que utiliza esta red social le gusta el tuit de un tercero, puede recuperarlo y publicarlo en su propia cuenta y ese “retuiteo” se toma como medida del interés que despierta un mensaje.

Cuanto más retuiteos reciba un posteo, mayor será su popularidad. En ocasiones, puede convertirse en tendencia (“*trending topic*”) y ubicarse entre los principales temas de conversación en Twitter.

Con el propósito de predecir en el menor tiempo posible qué tuits serán populares, Emanuel Meriles realizó su tesis de grado con el grupo de **Análisis y Procesamiento de Grandes Redes Sociales y Semánticas** de la Facultad de Matemática, Astronomía, Física y Computación (FaMAF), de la Universidad Nacional de Córdoba.

En 2017, ese equipo generó un modelo capaz de predecir cuándo un tuit sería repostado por un usuario particular, con una efectividad del 87%. El modelo ignoraba el contenido del posteo y solo se basaba en el comportamiento del usuario y sus contactos.

Un año más tarde, otra publicación del grupo demostró cómo las personas “influencers” son decisivas para definir si una publicación será o no popular, nuevamente de manera independiente de su contenido.

En este nuevo trabajo, realizado en 2020 por Emanuel Meriles, bajo la dirección de Martín Domínguez y Pablo Celayes, se incluyó la dimensión del tiempo: querían saber cuál es el menor lapso que debe considerarse para predecir de manera fiable si un tuit será tendencia.

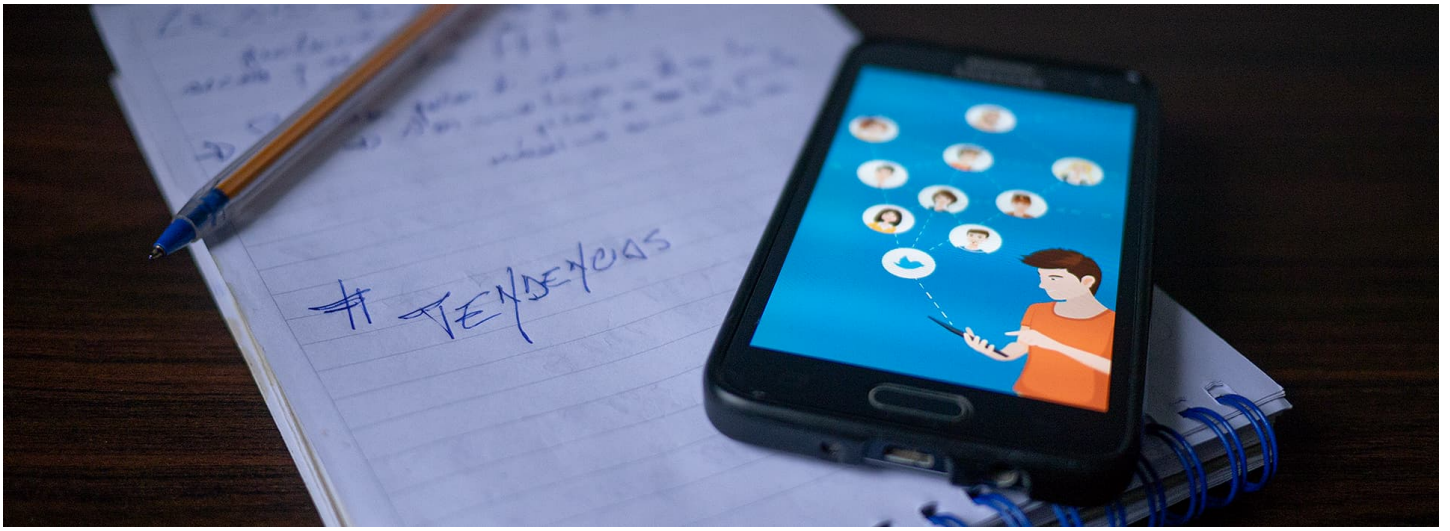
Para ello, analizaron 9,5 millones de tuits –publicados entre septiembre y octubre de 2018–, y trabajaron con un conjunto inicial de 5589 usuarias y usuarios, que finalmente quedó reducido a 3911 tras eliminar a quienes permanecían en inactividad.

Concluyeron que, si solo se analizan las conexiones sociales entre las personas usuarias, la primera hora de actividad de un tuit permite saber, con un 70% de efectividad, si ese posteo será popular. Si en cambio se evalúa el comportamiento de influencers, el tiempo necesario se extiende a 90 minutos. Finalmente, si también se analiza el contenido de la publicación, la misma performance se consigue antes, entre 30 y 60 minutos.

“El análisis de contenido hace referencia a un conjunto de técnicas de aprendizaje automático aplicadas al procesamiento del lenguaje natural (o sea el lenguaje coloquial). Lo que se hace es aplicar esas técnicas para hacer de palabras, oraciones o, como es nuestro caso, publicaciones de Twitter, vectores de números reales, para luego procesarlos matemáticamente o con fines predictivos”, explica Meriles.

En consonancia con publicaciones anteriores, el trabajo muestra que sigue siendo más importante la lista de contactos que el contenido del mensaje para predecir la popularidad de un tuit. Sin embargo, al combinar ambos insumos, los resultados se potencian.

“La importancia de este último trabajo es que ahora se puede saber cuánto tiempo se tiene que analizar un tuit para predecir si va a ser popular o no. Esto es muy importante para medir el impacto de ciertos mensajes y la diseminación de noticias. A partir de saber qué discursos se van a convertir en populares, podríamos anticiparnos haciendo contranarrativas para frenar el avance de información falsa, por ejemplo”, explica Domínguez.



Dos modelos predictivos

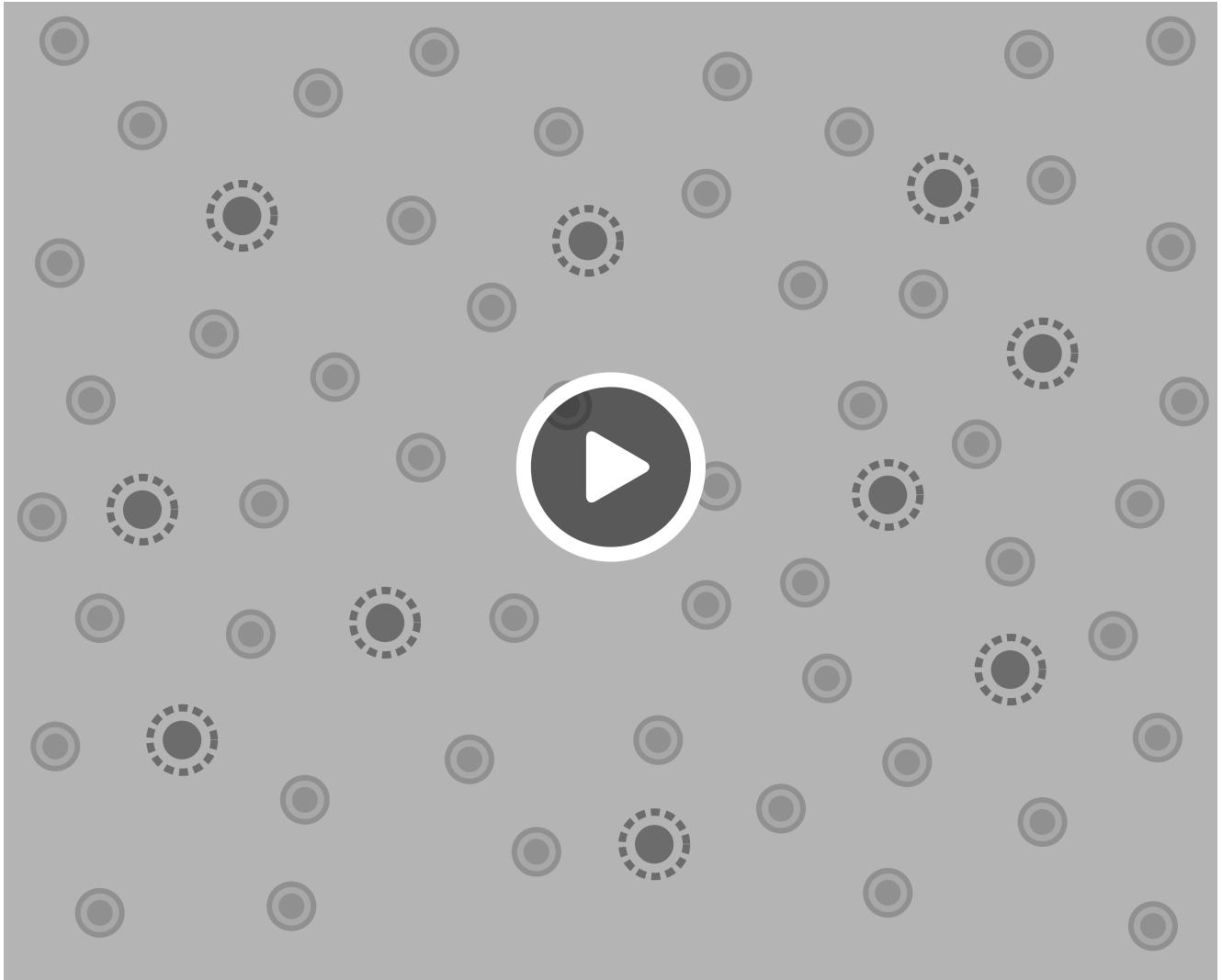


○ Usuaris/os de Twitter

● Usuaris/a/o central

○ Entorno de la usuaria/o central

El primer modelo estuvo destinado a predecir el comportamiento de un usuario o una usuaria central, cuyo entorno incluye al conjunto de personas seguidoras y seguidas. Cuando alguien del entorno del usuario o usuaria central publica un tweet, el modelo registra durante una hora los retuiteos que recibe de su entorno. Eso permite predecir con un 70% de certeza, si el usuario o la usuaria central también re-publicará el postero original.



○ Usuaría/o de Twitter ⊗ Influencer

En el segundo modelo, de todo el universo de personas usuarias, cada una fue clasificada como influencer o población general. Se eligió un tuit publicado o republicado por un usuario o usuaria general y se registró durante 90 minutos la actividad de retuiteo de influencer respecto a ese tuit. Así lograron predecir con un 70% de certeza si el tweet será popular entre los usuarios y las usuarias generales.

Sobre el grupo de Análisis y Procesamiento de Grandes Redes Sociales y Semánticas (APGSyS)

En su línea de investigación, aplican diferentes técnicas de aprendizaje y análisis, con el propósito de poder inferir comportamientos o realizar predicciones sobre redes semánticas y sociales.

En las redes sociales como Twitter, Facebook e Instagram, por ejemplo, la intención es poder predecir el interés de alguna persona usuaria en particular o el comportamiento masivo que determina la popularidad de una publicación.

En redes semánticas, como las derivadas de Wikipedia, intentan entender cómo evoluciona la información y poder predecir cuáles serían los cambios más probables basados en el desarrollo previo de dicha red.

Integrantes | Martín Ariel Domínguez, Damián Barsotti, Ezequiel Orbe y Pablo Ariel Duboue.

Para mayor información sobre APGSyS, visitar el [portal de la FaMAF](#).

La publicación

Twitter Early Prediction of preferences and tendencies based in neighborhood behavior, 2020.

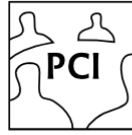


UNCiencia es la agencia de comunicación pública de la ciencia, el arte y la tecnología de la Universidad Nacional de Córdoba. Es una iniciativa de la Prosecretaría de Comunicación Institucional.

✉ unciencia@pci.unc.edu.ar | ☎ (0351) 5353730.



Universidad
Nacional
de Córdoba



Prosecretaría
de Comunicación
Institucional