

Software para Medir la Capacidad de Propagación de Usuarios de Twitter

Leonardo Ibáñez, Pablo Pellecchia, Martín Agüero

Grupo Turing, Facultad de Ingeniería, Universidad de Palermo.

Abstract. El estudio de usuarios de redes sociales es una actividad que ha surgido recientemente. Diferentes técnicas han sido desarrolladas para representar cuantitativamente la capacidad de un individuo para propagar ideas y opiniones por estas redes. La mayoría de estas técnicas sólo se enfocan en analizar las características intrínsecas de los mensajes.

En el presente trabajo, se realiza una introducción al tema y se analizan brevemente las características de las herramientas más populares. Se presenta un prototipo que, además de las técnicas mencionadas, utiliza el análisis semántico para obtener información relevante. El objetivo del proyecto se enfoca en distinguir aquellos usuarios de la red social *Twitter* más influyentes y efectivos para difundir ideas y opiniones a través de sus seguidores. Para medir el alcance de estos usuarios clave, se propone emplear un indicador denominado Índice de Propagación.

Keywords: social networks, twitter, semantic processing, graphs, software engineering.

1 Introducción

En la actualidad la masividad de las redes sociales está llevando a las personas a modificar sus conductas respecto de cómo se expresan y se informan [1]. Los usuarios de estas redes no solamente comparten lo que les sucede en sus vidas cotidianas a través de textos, fotos y videos, sino que también comentan sus sentimientos y opiniones acerca de la realidad que los rodea. Estos datos se han vuelto relevantes dentro del ámbito empresarial, académico y gubernamental para la elaboración de estudios sociológicos de la población. Los usuarios dentro de la red social interactúan compartiendo información y opiniones sobre diversos temas y, a su vez, producen temáticas nuevas que se esparcen a lo largo de la red. A pesar de la forma aparentemente caótica con que las interacciones entre los usuarios tienen lugar dentro de la red, ciertos temas compiten para conseguir una mayor atención y así llegar a la cima en términos de popularidad. Todo este fenómeno contribuye para formar nuevas tendencias y así crear la agenda pública de la comunidad [2].

Dentro de la gran cantidad de redes sociales existentes, el presente estudio ha seleccionado a Twitter [3] por su gran difusión y disponibilidad pública de datos. Twit-

ter es una red de *microblogging*¹ donde cada usuario mantiene un perfil, el cual está compuesto por su nombre, ubicación, lista de seguidores y cualquier otra información que se desee compartir. En este perfil se publican pequeños mensajes (*Tweets*), de no más de 140 caracteres, que a su vez actualizan los perfiles de los seguidores. Estos escuchan lo que publican sus seguidos y a su vez pueden responder a estas publicaciones usando herramientas, como los *retweets*, donde un usuario vuelve a publicar un tweet de otro usuario (equivalente al forward en el mail), *hashtags* (etiquetas que marcan el tema de una conversación) definidas como palabras clave y principales de un tweet y que siempre van antecedidas por un carácter “#”. Las herramientas antes nombradas y la utilización de mensajes cortos hacen que el procesamiento para la extracción de datos de Twitter sea más fácil al ser pública la relación entre usuarios y seguidores siendo así posible determinar cómo se propagó el tema a través de la red social.

En Twitter existen usuarios que adquieren mayor reputación por la manera en que influyen sobre su grupo de seguidores, ya sea porque son personas famosas fuera de Twitter (políticos, artistas, empresarios) o por la cantidad de personas que los siguen [4]. Estos usuarios se transforman en centros de propagación de información que sus seguidores pueden seleccionar y compartir. Se suele decir que estos usuarios establecen tendencia dentro de la red social, ya que son capaces de difundir temas que captan una mayor atención por el resto de usuarios. Twitter rastrea esos hashtags, utilizados con mayor frecuencia y los cataloga bajo el nombre de *Trending Topics*. Este factor ha sido aprovechado por las empresas que, por ejemplo, establecen campañas de marketing para el lanzamiento de un producto y desean observar cuál es la respuesta al mismo. En la última entrega de los premios Oscar, una importante compañía de dispositivos móviles publicó uno de sus nuevos productos a través de una fotografía que rápidamente se volvió Trending Topic, rompiendo el record de la mayor cantidad de retweets de la historia [5].

Una de las principales fortalezas de Twitter, y que se cita a menudo en la literatura, es su naturaleza de tiempo real, es decir, que los usuarios publican sobre eventos a medida que ocurren. Se afirma que actualmente Twitter difunde noticias antes que las cadenas informativas [6]. Por el contrario, muchas veces sucede que tendencias que se generan en Twitter influyen en acontecimientos por fuera de la red social. Un ejemplo reciente es el reciente levantamiento en Egipto, el cual fue en parte promovido a través de Twitter [7].

Los hechos mencionados han abierto nuevas investigaciones dentro de las ciencias de la computación, por ejemplo la minería social, donde se aplican conocimientos provenientes del procesamiento de lenguaje natural, minería de datos y recuperación de datos sobre redes sociales estudiando cómo se propaga la información [8]. En Weng et al. [9] se analiza cuáles usuarios son más influyentes que otros, un estudio similar ya había sido realizado en Agarwal et al. [10] para detectar autores de *blogs*² influyentes dentro de una comunidad de blogs. Esta investigación concluyó que los autores más influyentes no son necesariamente los más activos.

¹ Un servicio que permite a sus usuarios enviar y publicar mensajes breves, generalmente solo de texto.

² Un sitio web en el que uno o varios autores publican cronológicamente textos o artículos.

Existen sistemas más completos donde no solamente se analiza la influencia de un usuario sino que también realizan un análisis semántico a los mensajes que publica [11]. Finalmente, se pueden nombrar herramientas online como Hashtagify [12] o RetweetRank [13] que basan sus resultados en factores como los hashtags o los retweets de los usuarios únicamente y muestran los resultados a través de una interfaz gráfica Web.

En este artículo se presenta una novedosa herramienta, denominada TwitterDigger, para analizar la capacidad de propagación de las ideas/opiniones de usuarios de Twitter mediante el procesamiento de mensajes.

El proceso comienza con la recuperación de usuarios y tweets, manteniendo la relación seguidor - seguido. Luego analiza la relación que existe entre los tweets recuperados basándose en los hashtags, las menciones, los retweets, categorías y palabras clave mediante el análisis semántico de cada uno de los mensajes. También se propone un Índice de Propagación, para medir cuantitativamente la capacidad de un usuario de propagar una idea u opinión. Por último, la herramienta expone los resultados mediante una interfaz gráfica donde se representan visualmente con un grafo las relaciones entre usuarios donde se destaca la influencia entre seguido - seguidor.

2 Motivación

Dentro del ámbito de las redes sociales, han surgido una gran cantidad de estudios donde se analiza la influencia de los usuarios de la red social Twitter. Trabajos como el de Wu [8] evalúan un hecho puntual como la muerte de Michael Jackson, donde se monitoreó en tiempo real la manera en que la información se propagaba a través de la red social y la manera como los usuarios se influenciaban entre sí. En Wojciech et al. [14] se sugiere un modelo de predicción capaz de pronosticar cuándo una URL va a ser propagada por parte de un usuario en base a la probabilidad que tiene ese usuario de compartir esa URL. La probabilidad de que un usuario mencione una URL está basada en un análisis de la actividad anterior del mismo. También se puede destacar el trabajo de Cha et al. [4] donde se establece un nivel de influencia a través de la cantidad de seguidores, el número de retweets y la cantidad de menciones de un usuario. Una de las conclusiones de este trabajo es que no necesariamente tener una gran cantidad de seguidores asegura generar una gran cantidad de retweets o menciones en los seguidores. En Weng et al. [9] se propone un modelo para encontrar usuarios influyentes dentro de la red social utilizando una medida llamada TwitterRank que mide la influencia de un usuario tomando como referencia la similitud entre los tópicos que publica un usuario respecto de los de sus seguidores y la manera en que los usuarios están interconectados entre sí. Los trabajos antes mencionados estudian la influencia de un usuario sobre sus seguidores, pero ninguno lo hace realizando un análisis semántico de los tweets en estudio. En Chowdury et al. [15] se estudia a Twitter como un medio de propagación boca a boca. Se utilizaron como ejemplos productos y marcas y luego se analizó la estructura y sentimiento de los tweets que trataban alguno de estos factores. Otro estudio propuesto por los autores Go, Bhayani y Huang [11] propone un modelo para realizar un análisis semántico de los mensajes en Twitter extrayendo el sentimiento que expresan y clasificándolos en positivo o negativo, para esto

se utilizan algoritmos de aprendizaje automático. Este artículo dio lugar a Sentiment 140 [16] que es una herramienta Web para realizar el análisis de tweets en tiempo real.

En la actualidad hay disponible en Internet un gran número de herramientas online que analizan los datos de Twitter en tiempo real y presentan los resultados de manera visual. Se pueden citar como ejemplo a Twithacolic [17] y TrendsMap [18] que *rankean* a los usuarios y temas por zonas geográficas. Twitalyzer [19] le permite al usuario conocer su influencia sobre su grupo de seguidores en base al impacto de sus tweets. TweetReach [20] da como resultado que tan lejos se ha propagado el tweet dentro de la red, tomando como datos los retweets y quienes los recibieron. Otra herramienta es Kred [21] que nos muestra los contenidos que más influyen sobre la comunidad a la que pertenecemos para eso trabaja con dos índices, uno que muestra nuestra influencia sobre nuestros seguidores y el segundo nuestra actividad en Twitter. RetweetRank [13], Hashtagify [12] y SocialMention [22] estudian la influencia basándose únicamente en los hashtags, retweets y menciones de un usuario. Si bien existe una gran cantidad de sistemas disponibles en el mercado que permitan el estudio de la influencia de los usuarios en Twitter, son pocos los que combinan todas las variantes en que una idea se puede propagar a través de la red social y son menos los que realizan un análisis semántico de los tweets. Otro punto a tener presente es que no existen muchos sistemas que enfoquen su análisis al idioma español por lo que allí se observa una brecha sobre la cual proponer un sistema como TwitterDigger, siendo esta aplicación capaz de contemplar todas las variantes que se observa que las demás aplicaciones relevadas no satisfacen. A continuación se muestra una tabla comparativa con las principales herramientas relevadas (Ver Tabla 1).

Tabla 1 Comparación entre herramientas online de análisis de influencia

Herramienta	Característica	Tipo de análisis
Hashtagify	Utiliza el impacto de los hashtags para medir la influencia del usuario.	Hashtags
Sentiment140	Analiza tweets en tiempo real y aplica técnicas de análisis de sentimiento.	Análisis de sentimiento sobre tweets
SocialMention	Utiliza las menciones que se han hecho sobre un usuario para calcular su influencia.	Menciones
Twithacolic	Rankea los usuarios más populares en base a sus seguidores y a la cantidad de publicaciones.	Seguidores, tweets, retweets
Kred	Utiliza Hashtags y menciones y retweets para calcular un índice de influencia y medir la actividad en Twitter.	Hashtag, menciones, y retweets
TweetReach	Muestra que tan lejos se propagó un tweet dentro de la red social.	Retweets

3 Diseño del Prototipo

A continuación se describe la arquitectura general de TwitterDigger (Ver Fig. 1) que establece tres módulos principales: Recuperación de Datos, Análisis y Cálculo y finalmente el Módulo de Presentación de Resultados.

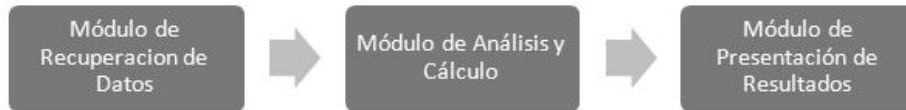


Fig. 1. Arquitectura general de TwitterDigger

3.1 Módulo de Recuperación de Datos

Este es el módulo (Ver Fig. 2) encargado de obtener datos (usuarios y mensajes) desde las fuentes que pueden ser online u off-line, una de esas fuentes es Twitter (recuperación y persistencia de datos *online*) y la segunda es una base de datos relacional (recuperación *off-line*) donde se persisten los datos de solicitudes anteriores. Con esos datos se construye un grafo de seguidores (*SocialGraph*) con una profundidad configurable a partir de un usuario específico (usuario raíz) que marca el punto desde el cual se desea comenzar con la recuperación de datos. En una posterior etapa se recuperaran los tweets y seguidores de estos usuarios.

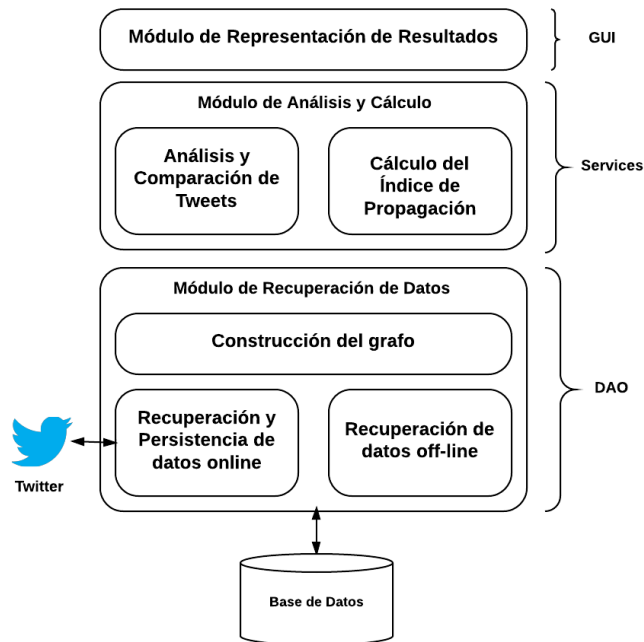


Fig. 2. Arquitectura detallada de TwitterDigger

Recuperación y persistencia de datos online.

Mediante la utilización de la API Twitter4J [23], TwitterDigger obtiene todos los datos desde Twitter. Para ello se diseñó un algoritmo de recuperación de datos que se configura a partir de un usuario raíz y un nivel de profundidad, definido por la cantidad de niveles de seguidores que se desean. El algoritmo recorre la red de seguidores del usuario raíz tomando los datos de éstos continuando luego con el siguiente nivel de seguidores hasta llegar a la profundidad deseada. Twitter limita la cantidad de solicitudes que se pueden realizar a sus servidores por sesión, para resolver este problema TwitterDigger conmuta entre 50 conexiones autorizadas evitando estas restricciones [24]. También en esta etapa se obtienen los últimos 200 tweets de cada usuario.

Una vez recuperados los datos desde Twitter, estos se persisten en una base de datos relacional para ser recuperados posteriormente.

Recuperación de datos off-line.

La recuperación off-line consiste en recuperar de una base de datos todos los usuarios y mensajes obtenidos en sesiones online anteriores.

Construcción del grafo.

Con los datos obtenidos tanto en el modo online como en el modo off-line se construye el SocialGraph (Ver Fig. 3), el cual es un grafo donde cada nodo representa a un usuario de Twitter, esta estructura contiene toda la información de cada usuario y sus últimos Tweets publicados. Las aristas o arcos que unen a los nodos representan la relación seguidor - seguido.

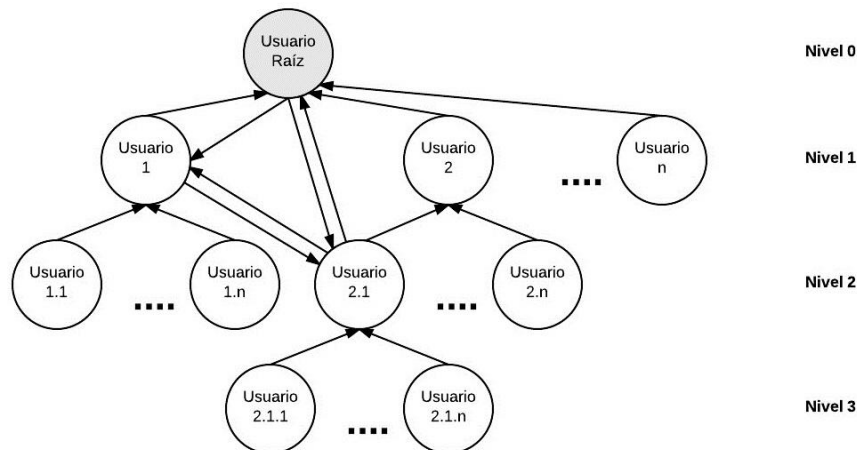


Fig. 3. Ejemplo de SocialGraph

Toda la información que se obtiene del módulo de recuperación de datos da como resultado una estructura con forma de grafo, tal como se observa en la Fig. 3. La es-

estructura obtenida servirá como entrada del módulo siguiente que analizará los datos contenidos y utilizará la información para el cálculo del índice de propagación.

3.2 Módulo de Análisis y Cálculo.

Este es el módulo (Ver Fig. 2) donde se comparan los tweets del usuario raíz frente a los tweets de todos los demás usuarios que componen el grafo. Para recorrer esta estructura se utiliza un algoritmo de búsqueda en amplitud [25] el cual explora horizontalmente todos los nodos del grafo buscando coincidencias de tipo Retweet (RT), Hashtag (HT), Categoría (CA) o Palabras clave (KW). El algoritmo de búsqueda verifica si un usuario fue analizado previamente para evitar entrar en ciclos y analizar un usuario más de una vez en caso de que este aparezca en más de un nivel del grafo. Para el análisis semántico de los tweets se utiliza la herramienta AlchemyAPI [26] que permite extraer las palabras clave y la categoría del texto que se procesa. Toda la información resultante, luego es utilizada para calcular el Índice de Propagación.

Análisis y comparación de tweets.

Con el algoritmo de búsqueda en amplitud se recorre el grafo en búsqueda de coincidencias entre los tweets del usuario raíz frente a los de los demás usuarios. Se considera coincidencia cuando, en los mensajes comparados, se observa alguna de las opciones descritas en la siguiente tabla (Ver Tabla 2).

Tabla 2 Tipos de coincidencia

Coincidencia	Análisis	Condición de coincidencia
Hashtag	Se analizan los hashtags usados por los seguidores.	Que posea exactamente los mismos hashtags.
Retweet	Se analizan los retweets hechos por los seguidores.	Que sea retweet de un mensaje emitido por el usuario raíz.
Categoría	Se extrae la categoría del texto del tweet.	Que coincida la categoría.
Keyword (palabras clave)	Se extraen palabras clave del texto del tweet.	Que la cantidad de palabras clave en común supere al 60% de las palabras clave del mensaje del usuario raíz.

En la Fig. 4 se observa un ejemplo de comparación semántica entre dos tweets que da como resultado: primero la coincidencia por categoría (CA) y luego también por palabras clave (KW). La comparación por palabras clave tiene que superar un umbral de coincidencia mayor a 60% (configurable) para considerar que 2 mensajes transmiten

el mismo concepto. En el ejemplo, el Tweet 2 posee un 66% de coincidencia de palabras clave.

Tweet 1:
Apple presentará el nuevo iPad el 22 de octubre en San Francisco

KW: Apple, nuevo Ipad, San Francisco
 CA: technology

Tweet 2:
 Un nuevo iPad será presentado por Apple la próxima semana

KW: nuevo Ipad, Apple
 CA: technology

Fig. 4. Ejemplo de comparación semántica

A continuación se explica el modo en que se calcula el Índice de Propagación en base a los datos recolectados en fases anteriores.

Cálculo del índice de propagación.

Luego de realizado el análisis de coincidencias se procede al cálculo del Índice de Propagación, el cual determinará, mediante un valor numérico, la capacidad del usuario raíz de propagar una idea hacia sus seguidores. En la sección 4 se detalla con mayor profundidad este tema.

3.3 Módulo de Representación de Resultados

Una vez calculado el Índice de Propagación se muestra por pantalla el resultado del procesamiento en forma de grafo. La interfaz de usuario es de tipo web y permite mostrar las relaciones e influencias de manera visual y simplificada.

Con tecnología *Servlet* y *JSP (JavaServer Pages)* se establece el nexo entre la GUI y el motor de *TwitterDigger*. Para representar gráficamente el *SocialGraph* se utiliza el proyecto *Cytoscape* [27] que es una librería de código abierto para la visualización y manipulación de redes. También se emplea la tecnología *JQuery* [28] y *JSTL (Java Server Pages Standard Tag Library)* [29] que ayudan a simplificar el código *JSP* y a encapsular las funcionalidades más utilizadas.

A continuación se presenta un ejemplo de la interfaz gráfica disponible para el usuario (Ver Fig. 5).

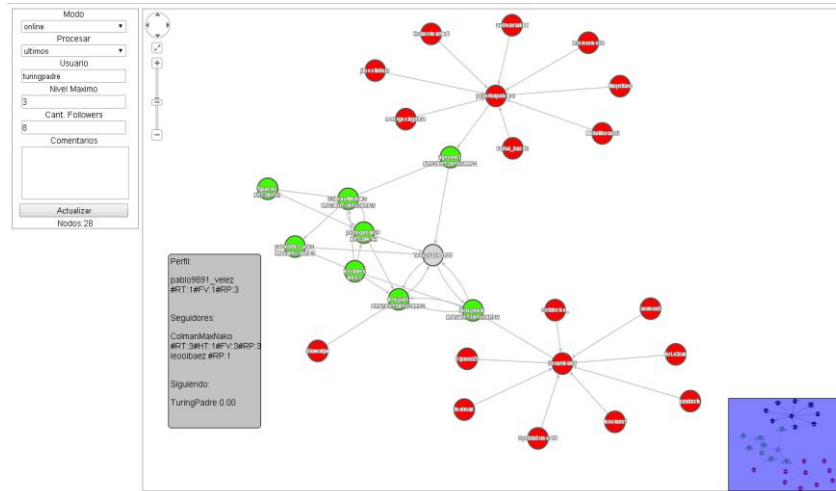


Fig. 5. Interfaz gráfica de usuario de Twitter Digger

El grafo generado representa a la red de usuarios seguidores de un nodo raíz. Ubicando el puntero del mouse, el sistema muestra un menú contextual donde, entre otras cosas, se muestra el Índice de Propagación asignado para ese usuario. En gris se muestra el usuario al cual se está analizando. En verde los usuarios en los que se encontraron coincidencias y en rojo los usuarios que no presentan coincidencias pero forman parte del análisis.

4 Cálculo de Índice de Propagación

El Índice de Propagación permite cuantificar la capacidad de un usuario de propagar una idea u opinión hacia sus seguidores. En los usuarios cuyo Índice de Propagación es más alto significa que tienen más chances de publicar una idea en su perfil de Twitter y que ésta sea tomada y publicada en los perfiles de sus seguidores.

Tabla 3 Factor por tipo de coincidencia

Tipo de coincidencia	Factor
Retweet	0.4
Hashtag	0.3
Keywords	0.2
Categoría	0.1

Para el cálculo de este índice se toma la información obtenida en la etapa de análisis y comparación de tweets dentro del Módulo de Análisis y Cálculo y en función del tipo de coincidencia se asigna un peso o factor para cada ocurrencia (ver Tabla 3). El

factor marca un grado de importancia a cada tipo de coincidencia por la relación que establece entre seguidor – seguido. Si una persona hace retweet, no solamente está involucrado con el contenido del mensaje sino que está marcando la fuente precisa desde la cual comparte la información. En cambio, si solamente se comparte un hashtag, se etiqueta el tema del que habla el tweet sin marcar la fuente desde la que se toma la idea. Es por eso que se ha decidido que el valor del factor del retweet sea más alto que el del hashtag.

A continuación se presentan la ecuación (1) que conforma al índice de propagación. Se define a s como la cantidad de seguidores de un usuario, a m la cantidad de tweets que emite un usuario y al Índice de Propagación con la siguiente ecuación:

$$IP = \frac{RTprop + HTprop + CAprop + KWprop}{m} \quad (1)$$

Donde $RTprop$ (2), $HTprop$ (3), $CAprop$ (4) y $KWprop$ (5) establecen la proporción de veces en que un usuario logró propagar una idea hacia sus seguidores y $cantRT$ (2), $cantHT$ (3), $cantCA$ (4) y $cantKW$ (5) muestran cuantas veces se encontró un tipo de coincidencia:

Proporción de Retweet:

$$RTprop = \frac{cantRT * fRT}{s} \quad (2)$$

Proporción de Hashtag:

$$HTprop = \frac{cantHT * fHT}{s} \quad (3)$$

Proporción de Categoría:

$$CAprop = \frac{cantCA * fCA}{s} \quad (4)$$

Proporción de Palabras Clave:

$$KWprop = \frac{cantKW * fKW}{s} \quad (5)$$

4.1 Caso de estudio

A continuación se presenta un caso donde se aplica el Índice de Propagación. En la

Referencias:

m : Cantidad de tweets publicados

s : Cantidad de seguidores

RTP: Proporción de Retweets

HTP: Proporción de Hashtags

CAP: Proporción de tweets con coincidencia de categoría
KWP: Proporción de tweets con coincidencia de Palabras Clave
IP: Índice de propagación

Tabla 4 se pueden observar los resultados parciales generados por el Módulo de Análisis y Cálculo. La columna "IP" corresponde al Índice de Propagación calculado para cada uno de los 5 usuarios analizados.

Referencias:

m: Cantidad de tweets publicados
s: Cantidad de seguidores
RTP: Proporción de Retweets
HTP: Proporción de Hashtags
CAP: Proporción de tweets con coincidencia de categoría
KWP: Proporción de tweets con coincidencia de Palabras Clave
IP: Índice de propagación

Tabla 4 Ejemplo caso de estudio

	m	s	Cant. RT	RTP	Cant. HT	HTP	Cant. CA	CAP	Cant. KW	KWP	IP
Usr 1	250	50	45	0.36	30	0.18	20	0.04	5	0.02	0.0024
Usr 2	10	423	284	0.2685	12	0.0085	9	0.0021	6	0.0028	0.0282
Usr 3	5	200	13	0.026	1	0.0015	1	0.0005	1	0.001	0.0058
Usr 4	20	100	82	0.328	24	0.072	8	0.008	6	0.012	0.021
Usr 5	5	356	120	0.1348	5	0.0042	36	0.010	19	0.0106	0.0319

A partir de los datos de la Tabla 4, se presenta el siguiente gráfico de barras (Ver Fig. 6).

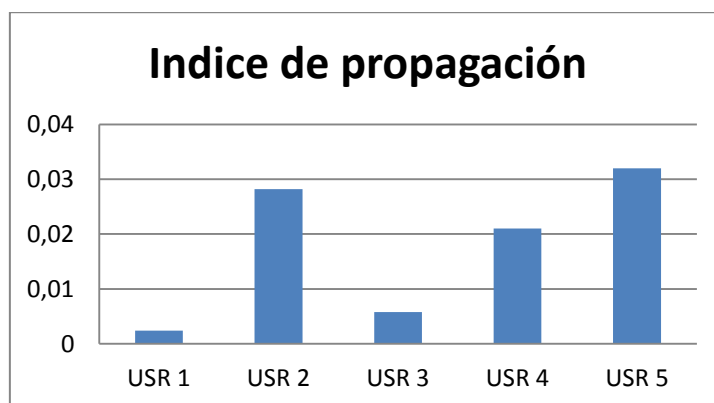


Fig. 6. Índice de Propagación comparado

El Índice de Propagación (IP) refleja el impacto de las publicaciones de un usuario sobre sus seguidores. En el caso del usuario 5 (USR 5), es quien posee el IP más alto debido a que con pocos tweets logra propagarse con mayor efectividad hacia sus seguidores. No sucede lo mismo con el usuario 1 (USR 1), quien publica una gran cantidad de mensajes sin lograr que éstos propaguen entre sus seguidores con tanta efectividad. Aunque, si bien su “RTP” es alto respecto de los demás usuarios, cuando la suma de las proporciones es dividida sobre la cantidad de publicaciones el índice de propagación disminuye.

5 Conclusiones y Trabajo Futuro

Se han presentado las herramientas de análisis de comportamiento de usuarios de la red social Twitter más relevantes junto con sus principales características. Es evidente que el análisis semántico de los mensajes aún es una técnica poco empleada para distinguir propagación de ideas y opiniones entre usuarios.

En función de la situación hallada, se propone una nueva herramienta que también emplee el análisis semántico como variable para determinar coincidencias entre mensajes. Asimismo, el estudio presenta un Índice de Propagación para describir y comparar cuantitativamente los usuarios más influyentes de la red social.

Para la próxima fase se proyecta diseñar casos de prueba donde se analizará a usuarios con miles de seguidores a fin de evaluar de forma intensiva tanto el software como el indicador propuesto.

El presente estudio no solo aborda el enfoque más difundido para analizar la influencia de usuarios de Twitter sino que también plantea al procesamiento semántico como una técnica válida para este fin.

6 Referencias

1. Kwak, H., Lee, C., Moon, S., Park, H.: What is Twitter, a Social Network or a News Media? En Proceedings of the 19th international conference on World wide web. (2010).
2. Asur, S., Galuba, W., Huberman, B. A., Romero, D. M.: Influence and passivity in social media. En Proceedings of the 20th international conference companion on World wide web, pp. 113–114. (2011).
3. Twitter, <https://twitter.com/>.
4. Benvenuto, F., Cha, M., Gummadi, K., Haddadi, H.: Measuring User Influence in Twitter: The Million Follower Fallacy. En Fourth International AAAI Conference on Weblogs and Social Media. (2010).
5. Infobae, <http://www.infobae.com/2014/03/05/1547977-cuanto-costo-la-selfie-mas-famosa-la-historia>.
6. Macdonald, C., McCreddie, R., Miles, O., Ounis, I., Petrovic, S., Shrimpton, L.: Can Twitter Replace Newswire for Breaking News?. En Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media. (2013).
7. El nuevo diario, <http://www.elnuevodiario.com.ni/nacionales/93355>.
8. Wu, F., Ye S.: Measuring message propagation and social influence on Twitter.com. En Lecture Notes in Computer Science, vol. 6430, Springer. (2010).
9. Jiang, J., Lim, E., Weng, J.: Twitterrank: Finding Topic-Sensitive Influential Twitterers. En ACM International Conference on Web Search and Data Mining. (2010).
10. Agarwal, N., Liu, H., Tang, L., Yu, P. S.: Identifying the influential bloggers in a community. En Proceedings of the international conference on Web search and web data mining, pp. 207 – 218. (2008).
11. Bhayani, R., Go Huang, L.: Twitter Sentiment Classification using Distant Supervision. Stanford University. (2009).
12. Hashtagify, <http://hashtagify.me/>.
13. RetweetRank, <http://www.retweetrnk.com/>.
14. Wojciech G., Karl A., Dipanjan C., Zoran D, Wolfgang K.: Outtweeting the Twitterers - Predicting Information Cascades. En Microblogs 3rd Workshop on Online Social Networks, WOSN. (2010)
15. Chowdury, A., Jansen, B., Sobel, K., Zhang, M.: Twitter power: Tweets as electronic word of mouth. En Journal of the American Society for Information Science and Technology. (2009)
16. Sentiment140, <http://www.sentiment140.com/>.
17. Twitaholic, <http://twitaholic.com/>.
18. TrendsMap, <http://trendsmap.com/>.
19. Twitalyzer, <http://www.twitalyzer.com/5/index.asp>.
20. TweetReach, <http://tweetreach.com/>.
21. Kred, <http://kred.com/>.
22. SocialMention, <http://www.socialmention.com/>.
23. Twitter4j, <http://twitter4j.org/en/index.html>.

24. Rate Limiting, <https://dev.twitter.com/docs/rate-limiting/1#search>.
25. Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C.: Introduction to algorithms (2nd edition). MIT Press and McGraw-Hill. (2001)
26. AlchemyAPI, <http://www.alchemyapi.com/>.
27. Cytoscape.js, <http://cytoscape.github.io/cytoscape.js/>.
28. JQuery, <http://jquery.com/>.
29. Jstl, <https://jstl.java.net/>.