

Efecto de los *trending topics* en el Volumen de Consultas a los Motores de Búsqueda

Santiago Ricci y Gabriel Tolosa
sricci.soft@gmail.com - tolosoft@unlu.edu.ar

Departamento de Ciencias Básicas, Universidad Nacional de Luján

Resumen Las redes sociales se han convertido en aplicaciones muy populares en Internet, principalmente para publicar información y comunicarse en grupos. Un caso bien conocido es Twitter, considerado un servicio de microblogging. Aquí se generan temas que se vuelven muy populares en la red social en un determinado momento, denominados *trending topics* (TT). En este trabajo se intenta determinar el efecto de estos temas en cuanto al volumen de consultas enviadas a un motor de búsqueda web. Para ello, se utiliza información de Twitter y tendencias de búsqueda. Los resultados iniciales muestran indicios de que los *trending topics* se utilizan luego para consultas al buscador: aproximadamente el 65% de las consultas muestran un aumento del interés cuando son TT e - inclusive - entre el 44% y 59% obtienen su pico de interés. Estos resultados se consideran indicios positivos respecto de la hipótesis planteada abriendo oportunidades de aprovechamiento de esta información para optimizar procesos internos de un motor de búsqueda.

Keywords: Twitter, trending topics, motores de búsqueda

1. Introducción

En los últimos cinco años las redes sociales se han convertido en aplicaciones muy populares, principalmente para publicar información y comunicarse en grupos de personas. Mientras que su crecimiento en cantidad de usuarios es exponencial, sus usos son de lo más variado (microblogging, content-sharing, perfiles profesionales, académicas, entre otras). Este tipo de sistema se basa principalmente en la existencia de conexiones “virtuales” entre sus participantes determinadas por el tipo de relación (amigo, seguidor, etc.) y que forman un grafo subyacente. Esta es una diferencia estructural respecto de los hyperlinks de la web donde los enlaces entre objetos son explícitos. La variación en la estructura afecta tanto la forma de obtener reputación en la red y cómo se localiza y disemina la información [10].

Las redes sociales permiten el agregado rápido de contenido y su propagación por el grafo. Uno de los casos bien conocidos es Twitter, una red social que cuenta con millones de usuarios alrededor del mundo [7]. Definida como un servicio de microblogging (ya que sus entradas, llamada *tweets*, tienen un máximo de

140 caracteres), entre sus características más notables se destacan los *trending topics* (TT) o tendencias. Estos resultan de un algoritmo que identifica los temas emergentes más populares¹. Los TT pueden ser términos, frases o *hashtags*² y se encuentran relacionados con los temas más populares en la red social en un determinado momento (por defecto, se determinan de forma personalizada para cada usuario). Es importante destacar, que en [7] se sugiere que gran parte de los TT están relacionados con las noticias del momento.

Por otro lado, se sabe que el buscador de Twitter es usado para monitorear cierto contenido, mientras que los motores de búsqueda Web son empleados para saber más acerca de dicho tema y que muchos usuarios ejecutan la misma consulta tanto en el motor de búsqueda de Twitter como en uno Web con el fin de capturar ambos usos [10]. También, es conocido el uso de Twitter para expresar opiniones acerca de diferentes temas, lo cual se ha traducido en gran cantidad de trabajos que plantean diferentes enfoques sobre cómo realizar minería de opinión sobre la red social [13] [14]. Estas premisas, junto a que según [7], gran parte de los usuarios de Twitter participan en *trending topics*, en conjunto permiten plantear la siguiente hipótesis: **“el hecho de que un tema sea trending topic, se traduce en un aumento en el volumen de consultas relacionadas con dicho tema en los motores de búsqueda Web”**. Si bien parece intuitiva su veracidad, no existen al momento propuestas de metodologías que permitan validarla ni cuantificar tal impacto.

En el presente trabajo se intenta obtener indicios que permitan verificar esta hipótesis. Esta cuestión es importante debido a que los motores de búsqueda deben responder millones de consultas (*queries*) por día, lo cual implica la necesidad de obtener eficiencia y efectividad para poder otorgar a los usuarios respuestas relevantes lo más rápido posible [11]. Entonces, si existe tal relación, puede sacarse provecho de la misma para mejorar el rendimiento a través del uso de técnicas conocidas en el ámbito de los motores de búsqueda por ejemplo, *caching* y *prefetching* de resultados.

Para poder validar definitivamente esta hipótesis se requiere contar con el conjunto de los TT en un periodo dado y la cantidad de consultas que recibió el motor de búsqueda respecto de estos temas en el mismo período. El primero de los conjuntos se puede obtener de forma directa usando la API de Twitter, pero el segundo es propiedad de los proveedores de los servicios de búsqueda y no se encuentra disponible. Para salvar esta situación se propone un método indirecto que – si bien no puede brindar datos absolutos – posibilita obtener indicios concretos acerca de la relación entre los TT y las consultas al motor de búsqueda. Como contribuciones principales, se propone un conjunto de métricas que, combinadas, permiten obtener una caracterización del comportamiento de un TT (mapeado en un *query*) en un período corto de tiempo. Además, se propone un método indirecto para contrastar el comportamiento de un conjunto de consultas antes, durante y después de ser *trending topic*. El uso de métodos indirectos

¹ <https://support.twitter.com/articles/101125-about-trending-topics>

² Cadena de texto precedida por el símbolo # que se utiliza como *keyword* de búsqueda.

es una técnica ampliamente utilizada en el ámbito de la recuperación de información distribuida [8] en casos que los proveedores de información no cooperen con el sistema (por ejemplo, *query-based sampling*). Complementariamente, se plantea un mecanismo para derivar consultas a partir de los TT.

Este artículo extiende trabajos previos [9] con resultados preliminares e incorpora el análisis una mayor cantidad de datos extraídos de la red social. El resto del trabajo se encuentra organizado de la siguiente manera: la sección 2 presenta trabajos relacionados con este estudio. La metodología utilizada, incluyendo las métricas propuestas y los datos analizados se introducen en la sección 3. Los experimentos y resultados se encuentran en la sección 4. Finalmente, se proponen cuestiones para discusión y trabajos futuros.

2. Trabajos relacionados

No es de conocimiento de los autores otros trabajos que traten la hipótesis planteada. Sin embargo, en [7] se realiza un estudio acerca de las características de los TT y se comparan las búsquedas populares ofrecidas por el servicio de Google Trends³ con los TT, presentando el grado de solapamiento entre ambos, el cual se encontró que es bajo. También se estudió la diferencia en la “frescura” (*freshness*) de los temas en ambos y se halló que en Twitter son más persistentes. Además, se concluye que los usuarios de la red social tienden a hablar sobre noticias y que gran parte (31 %) de los TT duran aproximadamente un día.

En el trabajo de Asur y otros [2] se estudian los TT y se afirma que aquellos con grandes duraciones están caracterizados por la naturaleza “resonante” del contenido de sus *tweets* asociados, el cual proviene, generalmente, de los medios de comunicación tradicionales. De este modo, Twitter se comporta como un amplificador selectivo del contenido generado por los medios tradicionales mediante cadenas de retweets. En [1], se utiliza a los TT como base para predecir los temas que se volverán populares en el futuro cercano. En [4] se estudia al servicio desde el punto de vista estructural y del contenido. Una publicación posterior [6] amplía dicha caracterización incluyendo la distribución geográfica. Aquí se identificaron diferentes clases de usuarios y su comportamiento, junto con patrones de crecimiento y tamaño de la red. Finalmente, en [10] se compara la tarea de búsqueda de los usuarios en Twitter respecto a los motores de búsqueda, pudiendo cuantificar algunas diferencias. Principalmente, hallaron que las consultas a Twitter son más cortas pero con palabras más largas y una sintaxis más específica. También se usan palabras comunes, se repiten más y cambian menos. Esto se debe a que los usuarios de Twitter habitualmente realizan búsquedas para monitorear nuevo contenido mientras que las búsquedas web se usan para conocer más sobre un tema. Además, los resultados entregados por ambos tipos de servicios son diferentes. Estas observaciones resultan interesantes para el proceso de mapeo entre TT y *queries*.

³ <http://www.google.com/trends/>

3. Metodología

El enfoque general se basa en un proceso de tres fases. Como se mencionó, no es posible obtener los datos de TT y queries para un mismo período de tiempo de forma directa. Para salvar la situación se proponen los siguientes pasos:

1. Capturar los *trending topics* de Twitter durante un periodo de tiempo
2. Derivar consultas a partir de los mismos (los TT pueden ser términos, frases o *hashtags* entonces es necesario mapearlos en un *query*)
3. Analizar la evolución de estos queries usando el servicio de Google Trends, aplicando una serie de métricas que intentan capturar su comportamiento.

3.1. Métricas

Para analizar la evolución de las consultas e intentar establecer indicios de que los TT influyen (aumentan) el volumen de consultas a un motor de búsqueda se proponen tres métricas que capturan aspectos diferentes:

1. **Variación del interés (*Var*):** es la variación porcentual del interés de la consulta derivada para un *trending topic*. Formalmente se define del siguiente modo: sea $I(n, q_{t_i})$ el interés en el *query* q asociado al *trending topic* t_i en el día n (día en que el tema se convierte en *trending topic*), entonces la variación porcentual respecto al día anterior ($n-1$) está dada por la siguiente ecuación:

$$Var(n, q_{t_i}) = 100 \frac{I(n, q_{t_i}) - I(n-1, q_{t_i})}{I(n-1, q_{t_i})} \quad (1)$$

En el caso que $I(n, q_{t_i}) = 0$ y $I(n-1, q_{t_i}) = 0$, $Var(n-1, q_{t_i}) = 0$; y para el caso que $I(n, q_{t_i}) \neq 0$ y $I(n-1, q_{t_i}) = 0$, $Var(n, q_{t_i}) = 100I(n, q_{t_i})$.

Esta métrica, intenta capturar el hecho de que si los TT influyen en el volumen de consultas, entonces debe existir una gran diferencia en el interés en dicha consulta entre cuando el tema es TT y cuando no.

2. **Cambio de tendencia (*T*):** es la cuantificación del cambio en la tendencia que experimenta cierta consulta derivada cuando el tema se convierte en TT, respecto a su tendencia en los siete días anteriores. Formalmente, si q_{t_i} es la consulta derivada para el *trending topic* t_i en el día n , $m_{q_{t_i}}(x, y)$ es la pendiente de la línea de tendencia para la consulta q_{t_i} entre los días x e y , entonces el cambio en la tendencia se define como:

$$T(q_{t_i}) = \frac{m_{q_{t_i}}(n-6, n) - m_{q_{t_i}}(n-7, n-1)}{|m_{q_{t_i}}(n-7, n-1)|} \quad (2)$$

En consecuencia debe cumplirse que, para una consulta que viene experimentando interés creciente, $T(q_{t_i})$ sea significativa para poder afirmar que el cambio fue consecuencia de que el tema sea *trending topic*.

3. **Detección de picos:** es la aplicación de un algoritmo de detección de picos (*burst detection*) como el presentado en [12]. Dicha publicación plantea que para descubrir regiones con valores “pico” en una serie temporal, debe calcularse la media móvil (MA) y tomar como valores pico aquellos que superen x desvíos estándar sobre el valor medio de MA. Por lo tanto, aquí se lo utilizará con el fin de plasmar que en el día que un tema es *trending topic*, debería registrarse un pico en el interés de su consulta asociada.

3.2. Datos utilizados

Para obtener los TT se utilizó la API REST de Twitter ⁴ y se consideraron solo los TT para Argentina (no personalizados). Se realizaron dos capturas. La primera, a la que llamaremos C_1 [9], fue realizada entre el 05/12/2012 y el 12/12/2012 (7 días). La segunda, (C_2), fue generada entre el 06/03/2013 y el 20/03/2013 (14 días). En ambas ocasiones, los TT fueron obtenidos a intervalos de 5 minutos aproximadamente (por límites impuestos por el proveedor del servicio). Bajo dichas condiciones, en el caso de C_1 se obtuvieron 2002 muestras de 10 TT. Eliminando los duplicados en todo el período de captura, se obtuvo un total de 573 TT y eliminando solo las repeticiones del mismo día se obtuvo un total de 727 TT. En el caso de C_2 , se capturaron 3954 muestras de 10 TT. Eliminando duplicados en todo el período de captura se obtuvieron 956 TT y quitando solo las repeticiones del mismo día se obtuvieron 1422 TT.

3.3. Derivación de consultas

Este aspecto es muy importante en lo que respecta a los objetivos del trabajo. Como se mencionó, la estructura de los TT no es similar a la de las consultas a un motor de búsqueda y para un *trending topic* se pueden derivar varios *queries*. Además, puede no quedar claro qué consulta puede derivarse o las resultantes pueden ser ambiguas (podría emplearse sus *tweets* correspondientes para intentar desambiguar). En el caso de los *hashtags*, puede ser difícil extraer los términos. También es importante tener en cuenta que, según [10], los *queries* realizados en el sistema de búsqueda de Twitter difieren en longitud, en función y en sintaxis respecto a los de un motor de búsqueda (ambos son *free-text queries*, pero en Twitter, '#' y '@' tienen significados especiales). En consecuencia, parece no ser adecuado emplear la consulta asociada a cada TT que devuelve la API de Twitter. Una cuestión importante, también aportada en dicha publicación, es que gran parte (45.95 %) del conjunto intersección entre los *queries* formulados en Twitter y en la barra de búsqueda de Bing son informacionales y acerca de celebridades.

Teniendo en cuenta lo anterior, se diseñó un procedimiento para derivar las consultas haciendo uso de los siguientes criterios: si el *trending topic* es una frase o un término, entonces la consulta es dicha frase o término. En el caso que sea un *hashtag*, se elimina el # y se intenta separar los términos haciendo uso de

⁴ <https://dev.twitter.com/docs/api>

las mayúsculas (*camelcase*) y los números. Si no se dispone de mayúsculas, se utiliza la primera sugerencia de Google (se consulta al motor de búsqueda) que normalmente obtienen los usuarios a medida que van escribiendo su consulta. Si no existen sugerencias, se ejecuta el TT en Google y se toma la búsqueda sugerida. En caso de no existir tal sugerencia, se toma como consulta el *hashtag* (sin el #).

3.4. Obtención de datos de tendencias de búsqueda

Para analizar los cambios en la popularidad de las consultas derivadas se requiere alguna fuente de información que disponga de los datos correspondientes a los queries derivados de los TT. Dado que no es posible tener acceso al log de consultas de un buscador web (menos aún en un período puntual), se optó por utilizar los datos del servicio Google Trends. Este servicio permite ejecutar una consulta y obtener la evolución del interés a lo largo del tiempo fijando una ventana temporal. Si bien la documentación⁵ de la herramienta no especifica el método exacto para calcular los valores que devuelve, se sabe que los datos reflejan el número de búsquedas de un término en comparación al total de búsquedas realizadas en Google a lo largo del tiempo⁶. Además, los datos están normalizados y son presentados en una escala de 0 a 100, donde el valor 0 se corresponde a la falta de datos (el servicio solo muestra los datos de los términos que sobrepasen un cierto límite en los volúmenes de búsqueda).

Luego, para cada *query* derivado de cada TT del conjunto de cada día, tanto de C_1 como de C_2 , se consultó la información empleando una ventana temporal de 30 días y el filtro de datos para Argentina⁷. Según la magnitud del volumen de búsquedas, la cantidad de puntos retornados por la herramienta puede variar. Para C_1 , 314 TT (43 % del total) obtuvieron 29 observaciones (puntos). En el caso de C_2 , el 55 % (788 TT) obtuvo 26 puntos. En ambos casos, la cantidad de observaciones retornadas por las consultas restantes no se consideran suficientes para los experimentos que aquí se proponen.

4. Experimentos & Resultados

4.1. Variación del Interés

Para cada TT se calculó la primera métrica (Ecuación 1). Luego, se calculó la cantidad de consultas que tuvieron interés creciente, decreciente y constante. La Tabla 1 resume los resultados obtenidos tanto para C_1 como para C_2 . De allí, es importante destacar, que en ambos casos más del 65 % muestra un interés creciente. En la Tabla 2 se muestra el detalle. Se puede apreciar que en aproximadamente la mitad de los casos, el crecimiento fue del 50 % o superior.

⁵ <http://support.google.com/trends/?hl=es#topic=13762>

⁶ http://support.google.com/trends/answer/87285?hl=es&ref_topic=13975

⁷ Para poder cumplir con los términos del uso del servicio se solicitó colaboración a un grupo de estudiantes entre quienes se repartió la tarea.

Tipo de variación	Porcentaje	
	C_1	C_2
Positivo	71,97 %	67,01 %
Negativo	19,43 %	27,92 %
Nulo	8,60 %	5,08 %

Tabla 1. Variación en el interés respecto al día anterior.

En la figura 1 se muestra el crecimiento acumulado para ambas capturas, las cuales presentan comportamientos similares: alrededor del 30 % del total de *trending topics* con interés respecto al día anterior creciente, obtuvo un aumento de más del 100 %. Sin embargo, pese a que un buen porcentaje de los TT obtiene un aumento significativo, hay que observar la tendencia que venían experimentando las consultas en una ventana temporal mayor. Si la misma es creciente y pronunciada, el hecho de que el tema sea tendencia no posee un gran efecto sobre la popularidad de la consulta.

Porcentaje de crecimiento (x)	Porcentaje con $Var(n, q_{t_i}) > x$	
	C_1	C_2
10	13,72 %	18,94 %
20	10,18 %	10,04 %
30	13,27 %	10,04 %
40	6,64 %	5,87 %
50	4,42 %	7,58 %
60	5,75 %	6,44 %
70	4,87 %	3,41 %
80	5,75 %	3,60 %
90	0,44 %	3,03 %
100	5,75 %	1,70 %
200	10,18 %	10,23 %
300	4,42 %	4,17 %
400	2,21 %	2,27 %
500	3,54 %	0,95 %
>500	8,86 %	11,73 %

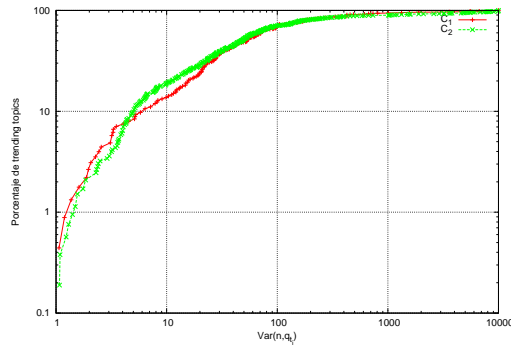


Tabla 2. Detalle del aumento del **Figura 1.** Función de distribución de la variación en el interés respecto al día anterior.

4.2. Cambio de tendencia de las consultas

Para cada consulta derivada se calculó la línea de tendencia de los 7 días anteriores a que el tema fuera TT (tanto en C_1 como en C_2). En la Tabla 3 se presentan los resultados. Para ambas capturas se observa que más del 50 % de los TT analizados posee una tendencia creciente. Esto plantea dos cuestiones: por un lado, pueden existir casos en los que la consulta derivada venía experimentando una tendencia decreciente, pero el interés aumentó al convertirse su tema asociado en TT. Por el otro, al menos que el cambio en la popularidad sea lo “suficientemente grande” no puede considerarse como válida la hipótesis. Por consiguiente, se elaboró la Figura 2 que relaciona a $Var(n, q_{t_i})$ con la pendiente de la línea de tendencia. El área recuadrada es la que debe observarse, dado que idealmente, la nube de puntos debería estar ubicada cerca del eje de las abscisas y alejada del eje de las ordenadas. Esto implica, que son de interés aquellas

consultas derivadas de los TT que venían experimentando popularidad relativamente constante hasta que el tema se convirtió en TT, momento en el cual se produce un aumento significativo en el interés. Obsérvese que en la figura, la nube de puntos está centrada en torno a $Var(n, q_{t_i}) \approx 80$ para ambas capturas y que la mayor densidad de puntos se encuentra en las pendientes con valores pertenecientes al intervalo $[-5; 5]$. Por lo tanto, existen casos donde se observa una tendencia decreciente en el interés y el día en que el tema se convirtió en TT, dicho interés creció abruptamente.

Tendencia	Porcentaje	
	C_1	C_2
Creciente	61,46 %	56,09 %
Decreciente	34,71 %	37,56 %
Estacionaria	3,82 %	6,35 %

Tabla 3. Tendencias de los 7 días previos a que un tema sea *trending topic*.

Observando específicamente las pendientes positivas, se encuentran casos en los cuales la variación porcentual de interés dada por la primer métrica es grande en comparación al crecimiento que debería esperarse dada la pendiente de la línea de tendencia y viceversa. Es importante aclarar que la tendencia que experimenta una consulta en los últimos 7 días no se corresponde necesariamente con la tendencia considerando una ventana temporal mayor. Por ejemplo, podría darse el caso de un *query* cuya popularidad evolucione de forma cíclica, con un período mayor a la ventana temporal actualmente considerada. El análisis de estos casos se deja para trabajos futuros.

Los resultados de la segunda métrica se muestran en la Tabla 4. Para C_1 , en aproximadamente en el 40% de los casos, el crecimiento de la pendiente es de al menos dos veces la pendiente de la recta que no considera el día en el que el tema fue *trending topic*. En cambio para C_2 , en el 33% de los casos aproximadamente $T(q_{t_i}) > 2$. Nuevamente, la ventana temporal empleada puede influir en los resultados. Si bien la figura no permite afirmar que la hipótesis se cumple para todos los casos, muestra que existen casos para los que pareciera que sí se cumple. Por otro lado, el análisis del cambio en la pendiente plasmado por la segunda métrica parece reforzar lo previamente dicho.

$T(q_{t_i})$	Porcentaje	
	C_1	C_2
< -0.5	5,31 %	12,12 %
< 0	3,98 %	6,63 %
< 0.5	11,50 %	14,20 %
< 1	15,49 %	15,34 %
< 1.5	12,83 %	8,52 %
< 2	10,62 %	10,04 %
< 2.5	4,87 %	3,60 %
< 3	3,98 %	3,41 %
< 5	9,29 %	7,95 %
< 10	9,73 %	9,47 %
≥ 10	12,39 %	8,71 %

Tabla 4. Relación entre pendientes.

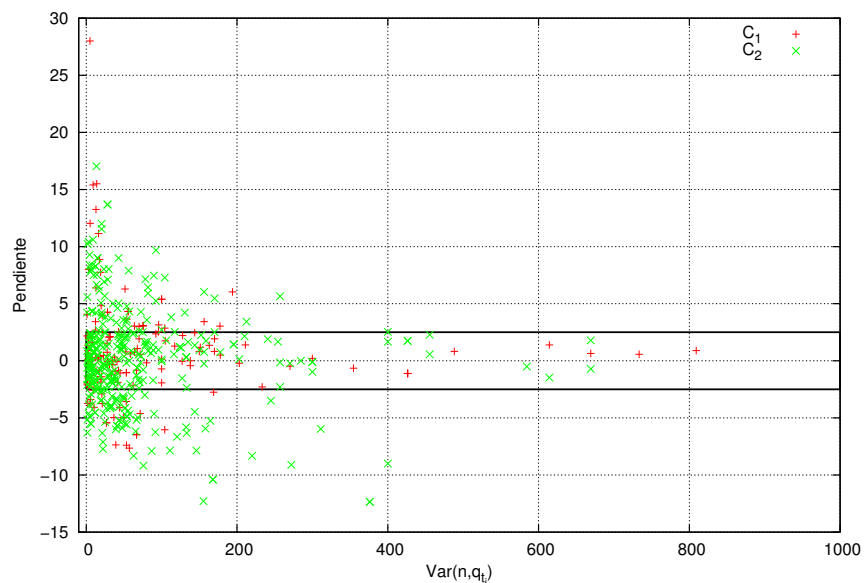


Figura 2. Relación entre la pendiente de la línea de tendencia del interés en la consulta en los 7 días anteriores a que el tema sea TT y la métrica determinada por la Ecuación 1.

4.3. Detección de picos

Por último se ejecutó el algoritmo de detección de picos sobre C_1 y C_2 , utilizando la configuración recomendada en [12], es decir, una ventana de 7 días y $x = 1,5$. En el caso de C_1 , del total de TT para los que se cuenta con suficientes datos, 162 (51,6 %) tuvieron un pico en el interés el día que fueron TT. En el caso de C_2 , 297 (37,69 %). La Figura 3 muestra el caso del *trending topic* “Papa Francisco”, del cual se derivó el *query* “papa francisco”. Esta consulta tuvo un pico en el interés el día en que su tema asociado fue *trending topic*. La Figura 4 muestra un caso donde no se identificó un pico de interés en la consulta. En ambas gráficas puede observarse la evolución de la media móvil hasta el día en que el tema fue *trending topic* y el umbral (*cutoff*) para considerar un pico en el interés.

Es curioso observar que, en el caso de la Figura 4, el pico se produce un día después que el tema fue *trending topic*. En realidad, el tema apareció como TT aproximadamente a las 22 hs. (22:09:38 es la hora exacta en la que aparece el primer registro), con lo cual se puede suponer que parte del tráfico va a estar contenido en el día siguiente. Por lo tanto, se procedió a analizar también el día siguiente. Los resultados muestran que la cantidad de consultas que obtienen un pico de interés se incrementa en un 7,6 % para C_1 y un 6,7 % para C_2 . Esto sugiere que debe estudiarse la duración de los TT. [7] y [9] presentan resultados de dicho estudio. Teniendo esto presente, se hace evidente la necesidad de estudiar

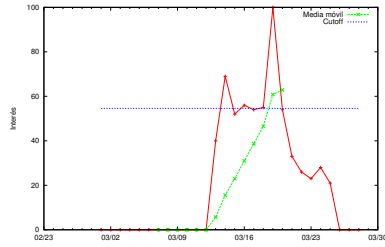


Figura 3. *Trending topic* de C_2 cuyo *query* obtuvo un pico en el interés.

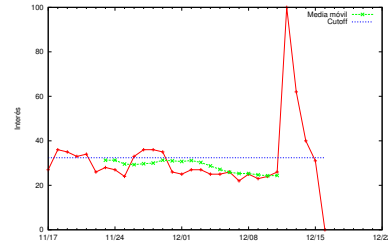


Figura 4. *Trending topic* de C_1 cuyo *query* no obtuvo un pico en el interés.

el aumento del tráfico en relación a la duración del TT, cuestión que se deja abierta para trabajos futuros. Esto es importante, dado que debe verificarse que la duración de un *trending topic* sea suficiente como para generar el tráfico necesario para alterar de forma abrupta la tendencia que venía experimentando la consulta.

5. Discusión y Trabajos Futuros

En este trabajo se intenta determinar el efecto de la aparición de información de tendencias en las redes sociales respecto de las consultas enviadas a un motor de búsqueda web. Para ello, se utilizan los *trending topics* de la red Twitter y la información de tendencias de búsqueda provista por Google como método indirecto, ya que no existe la posibilidad de obtener los archivos log de las consultas. Se proponen tres métricas que intentan capturar diferentes aspectos: variación porcentual, tendencia y picos. Los resultados presentan indicios de que los *trending topics* se utilizan luego para consultas al buscador. En particular, aproximadamente el 65% de las consultas muestran un aumento del interés (tendencia) cuando son TT e - inclusive - entre el 44% y 59% obtienen el pico de interés en ese mismo día o el siguiente. Estos resultados se consideran indicios positivos respecto de la hipótesis planteada aunque al momento y con los datos disponibles no se puede cuantificar.

Entonces, se abre la posibilidad de varios trabajos futuros: por un lado, seguir desarrollando métodos indirectos para obtener más información de otros sistemas similares y poder realizar mediciones con mayor grado de precisión. Esto incluye diferentes enfoques (por ej. semánticos) para obtener las consultas relacionadas a un TT. Además, debe estudiarse qué sucede luego de finalizado el período en el cual un tema es *trending topic*. Por otro lado, explotar esta relación para optimizar dos de los mecanismos principales utilizados por los servicios de búsqueda de gran escala como el caching y el prefetching de resultados. De esta manera, se podrían *anticipar* peticiones en momentos de menor actividad y disminuir las posibilidades de sobrecarga.

Referencias

1. Agarwal, S.; Agarwal, S. Social networks as Internet barometers for optimizing content delivery networks. In 3rd International Symposium on Advanced Networks and Telecommunication Systems (ANTS). 2009.
2. Asur, S.; Huberman, B. A.; Szabo, G.; Wang, C. Trends in social media: Persistence and decay. In 5th International AAAI Conference on Weblogs and Social Media. 2011.
3. Callan, J.; Connel, M. Query-based sampling of text databases. In ACM Transactions on Information Systems, v. 19, n. 2, pp. 97-130. 2001.
4. Java, A.; Song, X.; Finin, T.; Tseng B. Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07), pp. 56-65. 2007.
5. Jonassen, S.; Barla Cambazoglu B.; Silvestri F. Prefetching query results and its impact on search engines. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12), pp. 631-640. 2012.
6. Krishnamurthy, B.; Gill, P.; Arlitt M. A few chirps about twitter. In Proceedings of the first Workshop on Online Social Networks (WOSN '08), pp. 19-24. 2008.
7. Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World Wide Web (WWW '10), pp. 591-600. 2010.
8. Luo S. Federated search of text search engines in uncooperative environments. PhD Thesis. Carnegie Mellon University. 2006.
9. Ricci, S. Impacto de las Redes Sociales en la Popularidad de las Consultas a Motores de Búsquedas. Jornadas Argentinas de Informatica. JAIIO 2013.
10. Teevan, J.; Ramage, D.; Ringel Morris, M. #TwitterSearch: a comparison of microblog search and web search. In Proceedings of the fourth ACM international conference on Web Search and Data Mining (WSDM '11), pp. 35-44. 2011.
11. Tolosa G. H.; Feuerstein E. Mejoras algorítmicas y estructuras de datos para búsquedas altamente eficientes. En Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), p. 740-744. 978-950-766-082-5. 2012.
12. Vlachos, M.; Meek, C.; Vagena, Z.; Gunopulos, D. Identifying similarities, periodicities and bursts for online search queries. In Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data (SIGMOD '04), pp. 131-142. 2004.
13. Wang, X.; Wei, F.; Liu, X.; Zhou, M.; Zhang, M. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In Proceedings of the 20th ACM international conference on Information and Knowledge Management (CIKM '11) pp. 1031-1040. 2011.
14. Xinfan M.; Furu W.; Xiaohua L.; Ming Z.; Sujian L.; Houfeng W. Entity-centric topic-oriented opinion summarization in twitter. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD '12), pp. 379-387. 2012.