

Detección de patrones y grupos de sentimientos a partir del análisis de tuits políticos

Dra. Rocío Abascal-Mena

Universidad Autónoma Metropolitana - Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa Santa Fe,
05348, Ciudad de México, Distrito Federal, México, 5814-6502
mabascal@correo.cua.uam.mx

Dr. Erick López-Ornelas

Universidad Autónoma Metropolitana - Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa Santa Fe,
05348, Ciudad de México, Distrito Federal, México, 5814-6502
elopez@correo.cua.uam.mx

Dr. Sergio Zepeda-Hernández

Universidad Autónoma Metropolitana - Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa Santa Fe,
05348, Ciudad de México, Distrito Federal, México, 5814-6502
jzepeda@correo.cua.uam.mx

Resumen

Este artículo presenta un ejemplo de cómo pueden analizarse los tuits aplicando un léxico de opinión en dos distintos corpus conformados por tuits generados durante diferentes movimientos políticos y sociales en Francia. Ambos corpus corresponden a las mismas fechas con el propósito de llevar a cabo un análisis, durante el paso del tiempo, sobre las diferencias y similitudes entre los sentimientos. La investigación proporciona algunos elementos claves en los que se pueden apreciar *períodos y grupos de sentimientos* en una población en particular. El análisis efectuado a cada corpus es mostrado de manera gráfica de manera a poder visualizar patrones en el estado de ánimo de dos movimientos sociales diferentes. La metodología propuesta puede ser aplicada para la detección de importantes puntos de inflexión, incluso patrones de sentimientos, en mensajes enviados en manifestaciones sociales y políticas a través del uso de las redes sociales.

Palabras Claves: análisis de sentimientos, detección de similitudes, lingüística computacional, redes sociales, Twitter.

1. Introducción

El análisis de sentimientos es un área de investigación muy interesante que permite aplicar mediciones a los textos en línea para detectar, automáticamente, las emociones vertidas [1]. El auge de la Web social, en especial de Twitter, proporciona nuevas oportunidades para recopilar grandes cantidades de datos, en tiempo real, que son producidas directamente por los usuarios. Los grandes volúmenes de datos se pueden analizar de diferentes maneras con el fin de examinar patrones existentes en los mensajes enviados a través de las redes sociales. Los tuits pueden ser analizados para extraer las reacciones que tienen los usuarios ante diversos eventos y sucesos. Un aspecto importante de los tuits es que han sido, recientemente, utilizados como una forma de participación en movimientos sociales permitiendo la emisión de opiniones y reacciones ante diferentes sucesos. Estas opiniones están, por lo general, cargadas con sentimientos positivos y/o negativos hacia un movimiento o evento. Sin embargo, la apreciación de un evento y la manifestación por parte de los usuarios puede cambiar con el transcurso del tiempo al verse enfrentada a una serie de situaciones tales como: mayor información y/o el intercambio de mensajes con otros usuarios.

El estudio de Twitter, combinado con la participación política y social, es un campo de investigación que apenas comienza y que podría ayudar a entender las nuevas formas de comunicación que están mediadas por el uso de las llamadas Tecnologías de la Información y la Comunicación (TIC). Con la llegada de Twitter, la extracción de información actual y la conformación de corpus pueden realizarse de manera rápida. Por lo tanto, han habido una gran cantidad de investigaciones en lingüística computacional que han sido enfocadas al estudio de los tuits para detectar opiniones, sentimientos, cuestiones de género y racismo, etc.

El análisis de sentimientos a partir del uso de datos públicos reduce significativamente los costos, los esfuerzos y el tiempo necesario para administrar encuestas y cuestionarios [2]. Especialmente, se puede apreciar que Twitter es utilizado para extraer las preferencias políticas a partir del análisis del uso del humor contenido en los tuits. En [2] se utilizó POMS (en inglés: Profile of Mood States) para destilar, a partir de los mensajes de Twitter, distintas series de tiempo correspondientes a 6 diferentes atributos emocionales (tensión, depresión, cólera, vigor, fatiga y confusión). POMS es un instrumento psicométrico que proporciona una lista de adjetivos para los que el paciente tiene que indicar el nivel de aprobación. Cada adjetivo se relaciona con un estado de la mente y, por lo tanto, la lista puede ser aprovechada como base para un analizador de estados de ánimo contenidos en datos textuales. De igual manera, se hace uso de un léxico subjetivo para determinar el puntaje positivo y negativo para cada conjunto de datos correspondientes a un tuit [3]. En este caso, los números crudos de los tuits positivos y negativos, acerca de un tema determinado, se utilizan para calcular una puntuación de confianza (la relación entre el número de tuits positivos y negativos). Mediante una simple inspección manual de los tuits se pudieron encontrar ejemplos que fueron clasificados en un sentimiento incorrecto. Sin embargo, los autores utilizaron este método para medir algunos temas como la confianza del consumidor, la aprobación presidencial y la elección presidencial de 2008 en Estados Unidos.

Otro trabajo interesante en el área es el de [4] el cual contiene dos partes bien diferenciadas: en la primera LIWC (en inglés: Linguistic Inquiry and Word Count) se utilizó para realizar un análisis superficial de los tuits relacionados con los diferentes partidos políticos que compitieron en las elecciones federales de Alemania en 2009. En la segunda parte, los autores afirman que el solo recuento de los tuits, con referencias a uno de los partidos, refleja con exactitud los resultados de las elecciones.

Un creciente número de análisis empíricos de los sentimientos y estados de ánimo, polaridad, se basan en colecciones textuales de datos generados en Twitter en los cuales se aplican sofisticados algoritmos para procesar, aplicar reglas gramaticales y clasificar a los tuits en distintos estados de ánimo. Así, por ejemplo, existe la aplicación de clasificadores basados en un léxico [5, 6]. En [7] se presenta un clasificador que fue

desarrollado específicamente para el análisis de tuits a partir del uso de palabras claves, frases y emoticones para determinar el estado de ánimo de cada tuit. Asimismo, ya han sido propuestos varios métodos para el análisis de tuits con el fin de detectar los cambios de estado de ánimo de la gente durante todo el día [6, 8].

En el 2010, se presentó un estudio, efectuado durante tres años, en el que se midieron los cambios del estado de ánimo de la población de los Estados Unidos a partir del análisis de tuits políticos [9].

En general, los estudios que analizan los tuits mediante la combinación de diferentes algoritmos de análisis de emociones han sido capaces de dar nuevas pistas sobre el comportamiento humano como resultado [10, 11, 12, 13].

El campo de investigación en el análisis de sentimientos ha desarrollado métodos para detectar automáticamente el sentimiento que está contenido dentro del texto. Algunos de estos métodos se basan en el uso de un léxico, el enfoque léxico. De este modo, existen diferentes léxicos como lo es SentiWordNet [14] el cual atribuye puntajes de sentimiento con valores reales positivos y negativos para cada conjunto de sinónimos llamados *synsets* y que provienen de WordNet [15]. También, existe el léxico de subjetividad llamado MPQA (Multi-Perspective Question Answering) que se distribuye bajo la licencia pública GNU y que consiste en una lista de claves de subjetividad que fueron compiladas a partir de varias fuentes [16]. Otro léxico es el Harvard General Inquirer que añade información sintáctica, semántica y pragmática a partes de palabras que han sido etiquetadas a partir de un discurso [17]. El LIWC (Linguistic Inquiry and Word Counts) es una base de datos propietaria que consta de una gran cantidad de expresiones regulares que han sido categorizadas. Sus clasificaciones están muy correlacionadas con las de la Harvard General Inquirer.

Bing Liu [18] mantiene y distribuye libremente un léxico de sentimientos que consta de dos listas de palabras, en inglés, positivas (2006) y negativas (4783). Este léxico contiene variantes morfológicas, argot, y marcas sociales con el fin de ser capaz de analizar sentencias, incluso si son errores de ortografía.

Estos trabajos, son solo algunos ejemplos que, muestran que hay una gran ambición de desarrollar herramientas de minería de opinión para analizar los medios de comunicación social con el fin de poder distinguir y extraer lo que es importante e interesante [19]. En los últimos años, se han propuesto varios métodos para detectar automáticamente los sentimientos en los mensajes dejando evidencia de las limitaciones que todavía existen en el análisis de textos como lo es la desambiguación, el uso de la ironía y el sarcasmo, el entendimiento del contexto histórico y cultural, etc.

Considerando la importancia que tienen las redes sociales como medios de información y participación masiva, la hipótesis del presente trabajo se resume en:

H₀. Los sentimientos vertidos en las redes sociales tienden a semejarse en movimientos distintos.

La validación de la hipótesis se realizó a partir de un estudio de tuits en el que se aplicó el léxico de Bing Liu [18], para el análisis de la polaridad (sentimiento) de cada tuit. El objetivo fue detectar patrones en los sentimientos vertidos en las redes sociales, independientemente de la temática tratada en los tuits. Este ejemplo de aplicación, enfocado en tuits políticos, intenta contestar a la siguiente pregunta:

P₁. ¿Es posible detectar patrones de sentimientos en movimientos sociales distintos, en Twitter, a partir de la visualización y el estudio del comportamiento de sentimientos recuperados durante un periodo de tiempo?

Este artículo presenta una alternativa en el área de la lingüística computacional al combinarla con una interpretación gráfica en el análisis de sentimientos y lograr la visualización de patrones de estado de ánimo y grupos de tuits alrededor de una determinada polaridad.

2. Desarrollo

El campo de investigación destinado al análisis de los sentimientos, también conocido como la minería de opiniones, ha desarrollado muchos algoritmos para identificar si un texto en línea es subjetivo u objetivo, y si cualquier opinión expresada es negativa o positiva [20].

En la presente investigación se utilizó el enfoque léxico combinado con un análisis lingüístico con el fin de detectar los sentimientos, durante un período de tiempo, vertidos en los tuits sociales y políticos. El enfoque léxico comienza con una lista de palabras, el léxico de Bing Liu, en el cual éstas están precodificadas de acuerdo a su polaridad para ser posteriormente analizadas a partir de su número de ocurrencias dentro del corpus. El análisis lingüístico, en contraste, analiza la estructura gramatical del texto para predecir su polaridad, en conjunción con el léxico [1].

Los tuits que se extrajeron corresponden a dos movimientos políticos franceses y que fueron etiquetados, respectivamente, con *#cgt* y *#hollande*. En este caso, los tuits utilizados están escritos en francés, los cuales surgieron a raíz de un malestar de la sociedad francesa ante el actual Presidente François Hollande. Este malestar se hace, también, presente a partir de la Confederación General del Trabajo (CGT) la cual agrupa a los trabajadores sindicales. De esta manera, se seleccionaron los tuits que contenían como hashtag *#cgt* y *#hollande*. Para eso, se tradujo el léxico de Bing Liu al francés.

Gran parte del proceso fue realizado de manera manual para eliminar algunas palabras que en francés no expresan un sentimiento o que corresponden a algo diferente. Posteriormente, cada una de las palabras contenidas en un tuit se clasificó en “positiva” o “negativa” utilizando el léxico anterior. Así, cada tuit es analizado para contar el número total de palabras o términos “negativos” y “positivos” dando una cantidad total. Esta metodología no toma en cuenta el sarcasmo que transforma la polaridad de un enunciado aparentemente positivo o negativo en su contrario [21]. No obstante, mediante el análisis de un gran corpus (como lo es nuestro caso) los resultados finales muestran una aproximación con la realidad.

En la metodología propuesta, se analizan los tuits utilizando tres pasos (ver Fig. 1): en el primer paso (1), se extraen los tuits que contienen, respectivamente, el hashtag *#cgt* y

#hollande para conformar corpus diferentes. En la segunda etapa (2), se analizan los tuits extraídos para normalizarlos y crear cada uno de los corpus. En el último paso (3) se utiliza el léxico francés que cuenta, para cada tuit, el número de términos “buenos” y “malos”. Estos tres pasos se realizan de forma automática mediante el uso de R que es un lenguaje de programación interpretado diseñado para el análisis de datos estadísticos [22]. Se seleccionó R ya que contiene algunos paquetes (conjunto de funciones), como *twitteR* (para conectarse a Twitter y recuperar los tuits), que permiten realizar procesos de una manera más ágil y transparente.

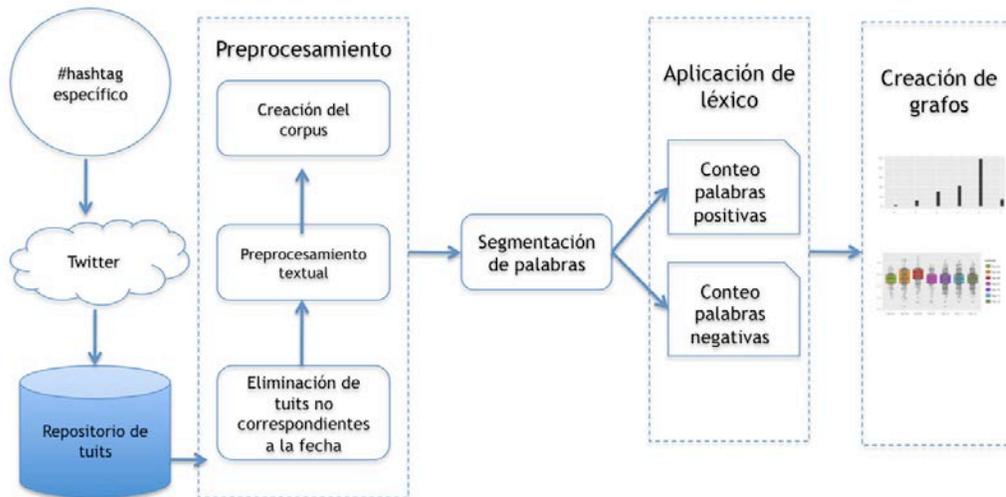


Fig. 1. Metodología aplicada desde la extracción de tuits hasta la creación de grafos para la detección de sentimientos y patrones en un corpus analizado.

A partir del uso de R se desarrolló un programa que es capaz de (ver Fig. 1):

1. Extraer tuits, automáticamente, que contienen un determinado hashtag (precedido por la almohadilla #), en este caso los tuits que contienen #cgt y #hollande. Es importante mencionar que el API de Twitter solo permite la recuperación de los tuits que tienen entre 6 y 9 días de creados. No es posible recuperar los tuits que tienen más de 10 días de haber sido creados, por lo que en algunos casos en los que la participación es muy alta, el proceso de extracción se llevó a cabo cada día.

2. Eliminar las palabras vacías y la puntuación.
3. Cambiar a minúsculas todas las palabras del tuit.
4. Extraer cada palabra de cada tuit para crear una lista de términos.
5. Buscar en el léxico francés cada una de las palabras que conforman la lista.
6. Contar el número de palabras positivas (+) y el número de palabras negativas (-) para obtener un número total para cada tuit. Este corresponde a la suma de las palabras positivas menos el total de las palabras negativas.
7. Crear un grafo de polaridad que muestra el número de palabras positivas, negativas y neutras (aquellos tuits que tienen 0 como cantidad total) para cada uno de los días que conforman el corpus.
8. Generar un grafo con la visualización de grupos de tuits alrededor de polaridades específicas.

Los pasos anteriores se realizaron para cada día de cada corpus. En algunos casos las etiquetas, #cgt y #hollande, son utilizadas simultáneamente por lo que los tuits que contienen las dos etiquetas forman parte de ambos corpus. Los dos corpus están compuestos por los tuits enviados entre el 4 de febrero y el 12 de febrero de 2014. Para cada corpus se eliminaron los días correspondientes al fin de semana (8 y 9 de febrero de 2014) debido a que hay una clara disminución del número de tuits enviados durante los fines de semana. El número de tuits, para cada día, que forma cada corpus se muestra en la Tabla 1.

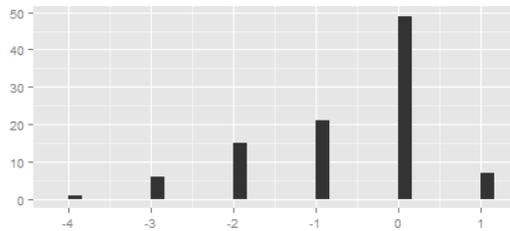
	FEBRERO 2014						
	4	5	6	7	10	11	12
#cgt	99	499	699	1198	1131	1152	599
#hollande	367	740	659	499	1836	1763	2000

Tabla 1. Número de tuits para cada día que conforma cada uno de los corpus.

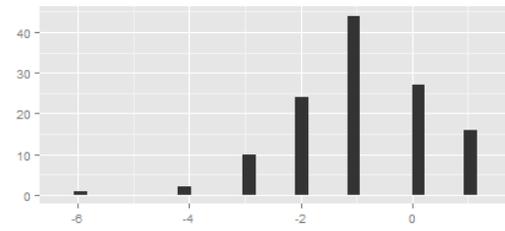
En la siguiente sección, se presentan los resultados obtenidos en la aplicación de la metodología, anteriormente descrita, a cada uno de los corpus generados.

3. Resultados

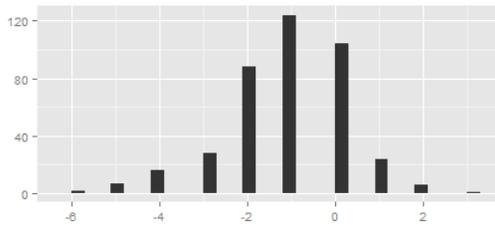
A partir de la aplicación del léxico de polaridad, traducido al francés se construyeron unos grafos que permiten visualizar ciertos “*patrones de emoción*” en los dos distintos corpus. Las siguientes tablas muestran el número de valores positivos, negativos y neutros obtenidos cada día para los corpus: #cgt (ver Tabla 2) y #hollande (ver Tabla 3).



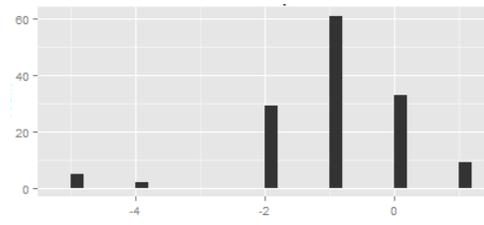
4 de febrero de 2014



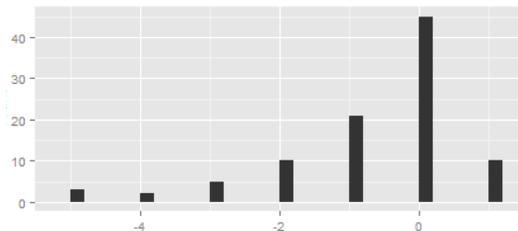
5 de febrero de 2014



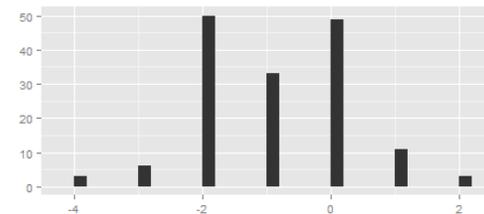
6 de febrero de 2014



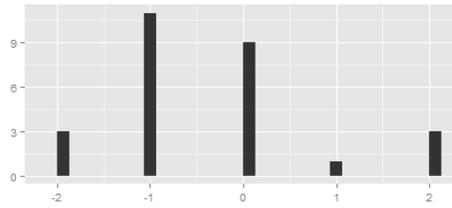
7 de febrero de 2014



10 de febrero de 2014

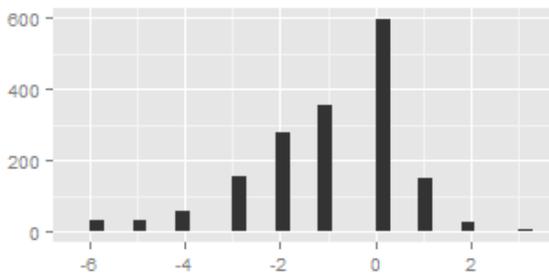


11 de febrero de 2014

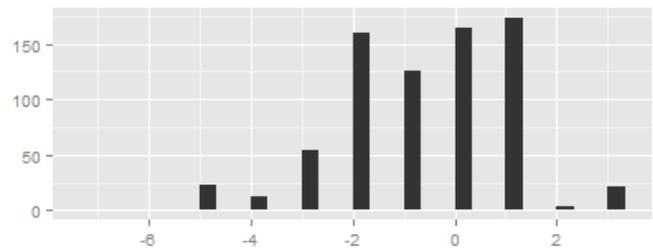


12 de febrero de 2014

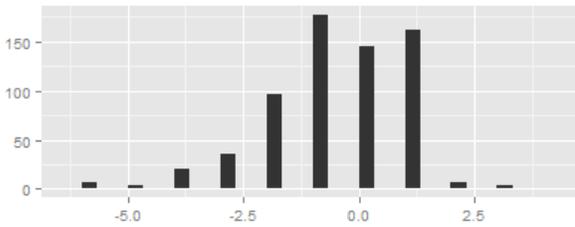
Tabla 2. Visualización de número de tuits (y) junto con la polaridad (x) para el corpus #cgt.



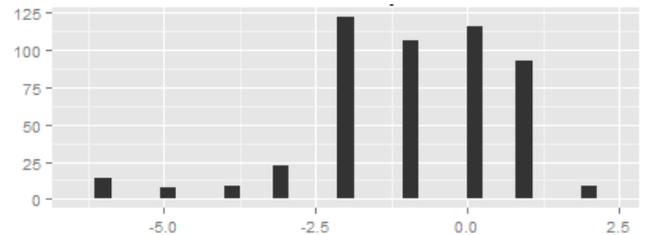
4 de febrero de 2014



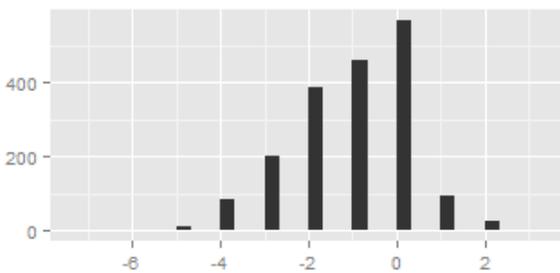
5 de febrero de 2014



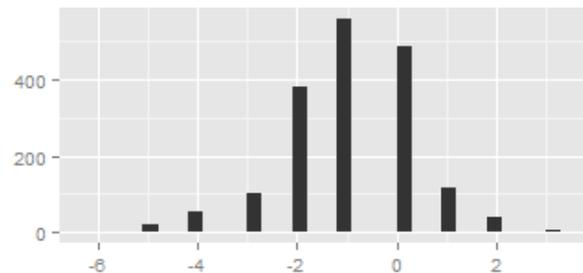
6 de febrero de 2014



7 de febrero de 2014



10 de febrero de 2014



11 de febrero de 2014

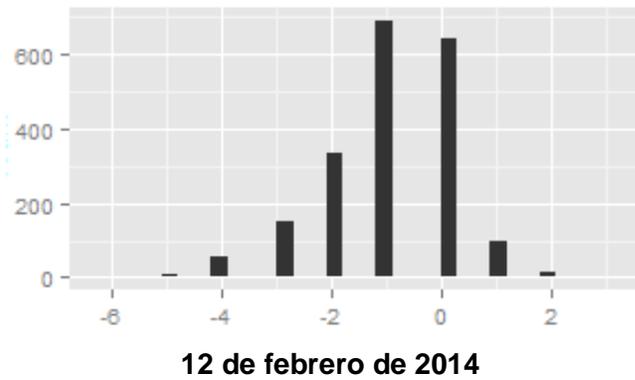


Tabla 3. Visualización de número de tuits (y) junto con la polaridad (x) para el corpus #hollande.

Identificación de patrones - El objetivo principal del presente trabajo, y que no se encontró en trabajos relacionados, tiene que ver con la posibilidad de identificar *patrones de emociones*, independientemente de que los temas políticos tratados en los corpus sean diferentes.

A partir de la visualización de las fluctuaciones que tienen los tuits en sentimientos positivos y/o negativos, a lo largo del tiempo, es posible hacer una comparación para corpus distintos. En la siguiente imagen (ver Fig. 2) se muestra el grafo concerniente al movimiento #cgt y que se analizó para los días 4 al 12 de febrero de 2014 (excluyendo el fin de semana del 8 y 9 de febrero). En este puede observarse un patrón que se repite en algunos días: *los extremos son muy bajos y la concentración se encuentra en el centro del grafo*. En este caso, por ejemplo, para los días 7, 10, 11 y 12 de febrero los extremos son (-5, 3). Es decir, que el máximo puntaje alcanzado para un tuit negativo es -5 y el más positivo tiene un puntaje total de 3. Para los días anteriores, 5 y 6 de febrero, los extremos son (-3, 2). Los cambios que se notan entre los grafos del 5, 6 y 7 de febrero son debidos, únicamente, a la cantidad de tuits analizados. Esto ocurre, también, para los siguientes días (10, 11 y 12 de febrero) donde los grafos obtenidos son idénticos. La concentración máxima de tuits durante los primeros días tiene una puntuación de 1

(positiva). La puntuación -1 (negativa) se obtuvo en los últimos días del estudio. Al analizar este grafo (ver Fig. 2) se puede llegar a dos puntos:

- En los primeros días la gente está más entusiasta pero como van pasando los días los tuits se van construyendo a partir de un mayor uso de palabras negativas. Es decir que los mensajes positivos tienden a volverse negativos a medida que pasa el tiempo. El contagio de sentimientos negativos es superior al de mensajes positivos.
- El grafo (el movimiento social) tiende a estabilizarse al mismo tiempo que muestra dos patrones diferentes. En los primeros días los valores obtenidos para un tuit son (-5, 1, 3). Para los últimos días los valores son: (-3, -1, 2). Estos valores representan el extremo inferior (puntuación negativa), la concentración máxima de tuits, y el extremo superior (puntuación positiva).

La Fig. 2 muestra que en los primeros días del movimiento #cgt es posible ver una diversidad de resultados. Por ejemplo, para el 5 de febrero se puede observar que los tuits se distribuyen en -3, -2, -1, 1 y 2 puntos. Este es el mismo caso para los próximos días donde la cantidad de tuits es casi la misma en diferentes puntuaciones tanto positivas como negativas.

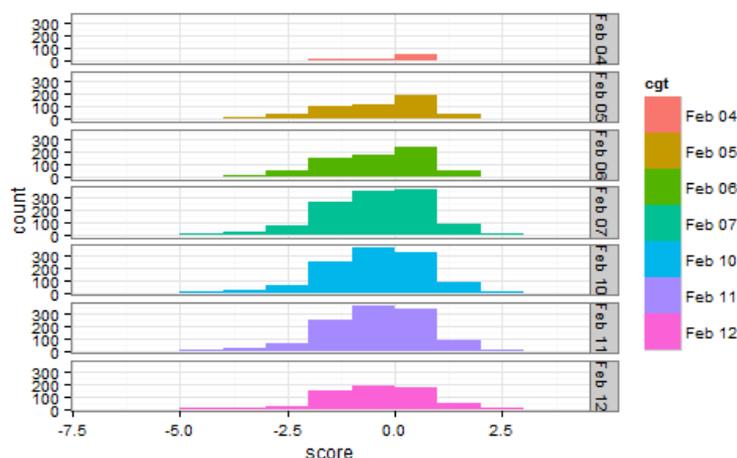


Fig. 2. Visualización de puntuación y número de tuits analizados cada día para el corpus #cgt.

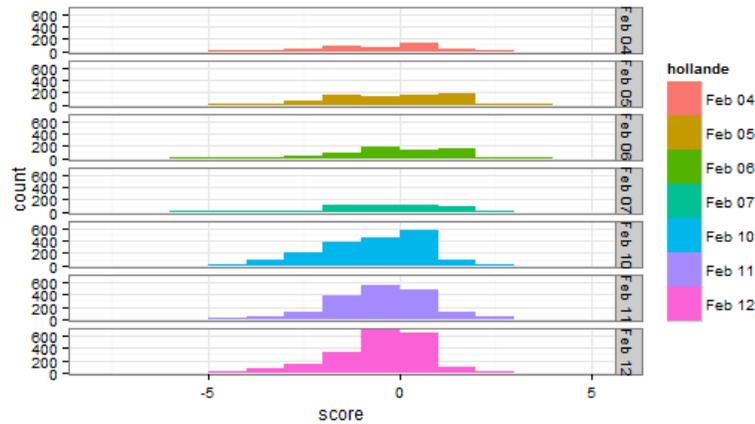


Fig. 3. Visualización de puntuación y número de tuits analizados cada día para el corpus #hollande.

Como van pasando los días, se puede ver que el grafo tiende a equilibrarse ya que hay una puntuación (en medio) que divide al grafo en dos partes iguales. Este es el caso, por ejemplo, para el 11 de febrero en el corpus de #hollande (ver Fig. 3) pero también se puede observar este mismo equilibrio para los días 10, 11 y 12 del corpus #cgt (ver Fig. 2).

Tomando en cuenta las características de los dos grafos anteriores (ver Fig. 2 y Fig. 3) es posible compararlos con los procesos de transformación social, en los cuales se tiene: *una fase de aceleración, una de estabilización y finalmente una de degeneración*. En todas estas fases los cambios estructurales se llevan a cabo a través de una acumulación de mensajes culturales y socio-políticos que influyen entre sí.

Comparando los resultados de #cgt y #hollande, como van pasando los días, se pueden observar similitudes entre sus grafos: las puntuaciones se centran entre -2.5 y 2 puntos. Al seleccionar el valor 0 (neutro) como la mitad de la gráfica se puede concluir que:

- en los primeros días del análisis, una puntuación positiva tiene más tuits que cualquier otra puntuación negativa (véase por ejemplo el grafo de #hollande para el día 4 y 5 de febrero);
- en los últimos días del análisis, la mayor cantidad de tuits obtiene una puntuación negativa;
- independientemente del número de tuits analizados, como van pasando los días el grafo tiende a equilibrarse (como una pirámide). En este caso, ya no existen fluctuaciones que hagan variar la forma del grafo.

4. Discusión

El presente trabajo aporta dos aspectos interesantes que son complementarios a un análisis lingüístico con técnicas ya conocidas (como las presentadas en la sección 1):

1. La capacidad de identificar comportamientos específicos alrededor de las discusiones sociales en redes sociales. Es decir, la posibilidad de identificar patrones de conducta alrededor de términos y sentimientos durante una determinada discusión.
2. La capacidad de visualizar grupos de tuits en torno a una determinada puntuación, es decir, que corresponden a un sentimiento en particular.

El movimiento #cgt muestra una explosión de tuits entre el primero y el segundo día analizado. Sin embargo, se puede observar una acumulación en torno a la mediana y el incremento de tuits en las puntuaciones negativas. Además no hay un gran cambio, en los días consecutivos, en términos de los tuits positivos ya que estos tienden a estabilizarse. Es decir que, a medida que pasa el tiempo los tuits tienden a volverse más negativos.

En el caso de #hollande se puede apreciar que la mediana es la misma, con un valor de 1, para todos los días que se analizaron. Por ejemplo para el 10, 11 y 12 de febrero existe un grupo de tuits concentrados en la mediana, pero otra en el cuartil superior (los tuits positivos) y otra en el cuartil inferior (los tuits negativos). Es posible ver este mismo

aspecto, el 5, 6 y 7 de febrero pero con grupos que tienen una concentración menor de tuits.

Comparando los dos diagramas de caja (ver Fig. 4 y Fig. 5) de #cgt y #hollande se puede observar que los tuits están muy dispersos en los primeros días. Como van pasando los días, los tuits tienden a agruparse y concentrarse en grupos haciendo que haya un mayor número de tuits en puntajes específicos. De esta manera, es posible identificar 5 grupos de tuits en nuestro análisis de ambos corpus.

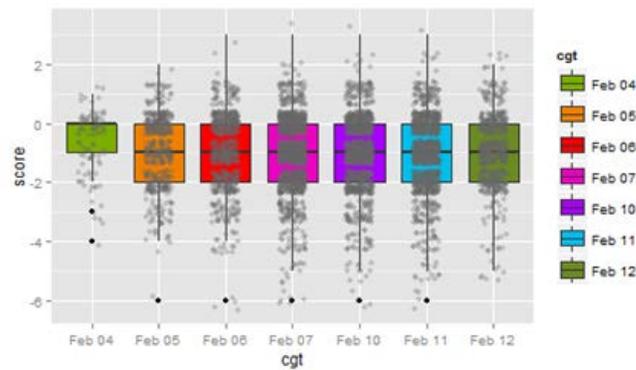


Fig. 4. Grupos de tuits, correspondientes al corpus #cgt, alrededor de una determinada polaridad.

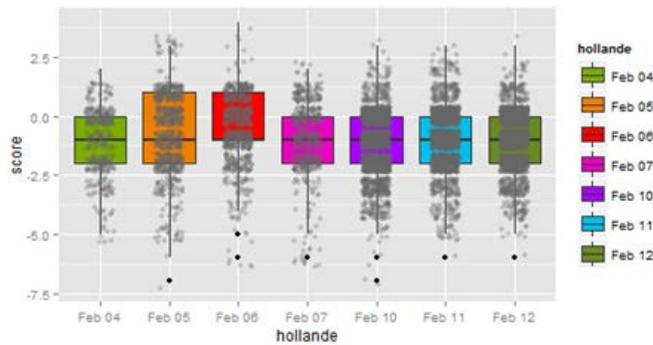


Fig. 5. Grupos de tuits, correspondientes al corpus #hollande, alrededor de una determinada polaridad.

5. Conclusiones

En este artículo presenta una metodología que es complementaria a la descrita en [23] en la que se estudia la evolución de un corpus de tuits a partir de la aplicación de la teoría de grafos. En la metodología, aquí descrita, se utiliza un léxico en francés que ha sido anotado en base a la polaridad de los términos.

El objetivo principal de la presente aplicación es la identificación de patrones de sentimientos, durante un periodo de tiempo, al comparar dos corpus formados por tuits políticos. El sistema propuesto realiza, de manera automática, la extracción, normalización, anotación y cálculo de la polaridad de cada tuit. Los grafos resultantes muestran que existe una semejanza en la conducta (sentimientos emitidos) de los usuarios cuando emiten sus tuits en un periodo de tiempo. Esta misma metodología puede ser aplicada a cualquier idioma siempre y cuando se haga una revisión manual de términos que son propios a la cultura.

A partir de esta metodología es posible identificar y visualizar grupos de tuits, así como su evolución durante el tiempo, dentro de la tabla de polaridad. Asimismo, es posible identificar a los actores y hacer un seguimiento de su posición política a lo largo de un movimiento social y político mediado por las redes sociales.

El trabajo futuro se centra en el estudio de la comparación de la metodología con otros idiomas. De igual manera, el análisis de tuits con una polaridad específica puede ser aplicado en la extracción de mensajes y actores claves que tienen, generalmente, una función de persuasión y que terminan por convencer al resto de la ciudadanía. Este es un trabajo interdisciplinario que implica no solo la aplicación de la computación sino además áreas cercanas a la sociología, las ciencias sociales y las humanidades. El estudio de los tuits en contextos de importancia social son claves para entender la evolución, comportamiento e involucramiento de la sociedad a través del uso de las redes sociales.

6. Referencias

- [1] M. Thelwall, K. Buckley, G. Paltoglou, "Sentiment in Twitter events". *Journal of the American Society for Information Science and Technology*, 62(2), 2011, pp. 406-418.
- [2] J. Bollen, H. Mao, Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 2011.
- [3] B. O'Connor, R. Balasubramanyan, B. R. Routledge, N. A. Smith, From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 2010.
- [4] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welp, Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 2010. pp. 178-185.
- [5] V. Wijaya, A. Erwin, M. Galinium, W. Muliady, Automatic mood classification of Indonesian tweets using linguistic approach. *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*. 2013. pp. 41-46.
- [6] V. Martínez, V. M. González, Sentiment Characterization of an Urban Environment via Twitter. In *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*. 2013. pp. 394-397. Springer International Publishing.
- [7] K. Z. Bertrand, M. Bialik, K. Virdee, A. Gros, Y. Bar-Yam, Sentiment in New York City: A High Resolution Spatial and Temporal View. *arXiv preprint arXiv: 1308.5010*. 2013.

- [8] V. Lampos, T. Lansdall-Welfare, R. Araya, N. Cristianini, Analysing mood patterns in the United Kingdom through twitter content. arXiv preprint arXiv:1304.5507. 2013.
- [9] M. Cha, H. Haddadi, F. Benevenuto, K. P. Gummadi, Measuring User Influence in Twitter: The Million Follower Fallacy. Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM), Washington, DC, 2010.
- [10] J. E. Chung, E. Mustafaraj, Can collective sentiment expressed on Twitter predict political elections? W. Burgard & D. Roth (Eds.), Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2011). 2011. pp. 1768-1769. Menlo Park, CA: AAAI Press.
- [11] P. S. Dodds, C. M. Danforth, "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents". Journal of Happiness Studies, 11(4). 2010. pp. 441-456.
- [12] A. Gruzd, S. Doiron, P. Mai, Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. Proceedings of the 44th Hawaii International Conference on System Sciences. Washington, DC: IEEE Computer Society. 2011.
- [13] A. D. I. Kramer, An unobtrusive behavioral model of "Gross National Happiness". Proceedings of CHI 2010. 2010. pp. 287-290. New York: ACM Press.
- [14] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010. pp. 2200-2204. European Language Resources Association.
- [15] C. Fellbaum, "WordNet: An Electronic Database. Cambridge". MA: MIT Press. 1998.
- [16] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the Conference on Empirical Methods in Natural Language Processing. ACL. 2005.

- [17] P. J. Stone, C. D. Dexter, S. S. Marshall, D. Ogilvie, "The General Inquirer: A Computer Approach to Content Analysis". Cambridge, MA: MIT Press. 1966.
- [18] L. Bing, "Sentiment Analysis and Opinion Mining". Morgan & Claypool Publishers, 2012.
- [19] D. Maynard, K. Bontcheva, D. Rout, Challenges in developing opinion mining tools for social media. Proceedings of @ NLP can u tag# user_generated_content, 2012.
- [20] B. Pang, L. Lee, Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 1(1-2), 2008. pp. 1-135.
- [21] R. González-Ibáñez, S. Muresan, N. Wacholder, Identifying sarcasm in Twitter: a closer look. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Short papers-Volume 2. Association for Computational Linguistics, 2011. pp. 581-586.
- [22] The R Project for Statistical Computing. <http://www.r-project.org/>. Abril 2014.
- [23] R. Abascal-Mena, R. Lema, F. Sèdes, From Tweet to Graph: Social Network Analysis for Semantic Information Extraction. IEEE Eighth International Conference on Research Challenges in Information Science. Marrakesh, Morocco. ISBN: 978-1-4799-2393-9. 2014. pp. 227-236.

7. Autores

La Dra. Rocío Abascal-Mena es profesor-investigador en el Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Realizó su doctorado en el Instituto Nacional de Ciencias Aplicadas (INSA) de Lyon, Francia, en el 2005. Sus intereses actuales de investigación incluyen: web semántica, lingüística computacional, recuperación y análisis de información, Interacción Humano Computadora, análisis de redes sociales.

El Dr. Erick López-Ornelas es profesor-investigador en el Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Recibió su doctorado en Informática en la Universidad Paul Sabatier de Toulouse Francia, en el 2005. Sus intereses de investigación incluyen: sistemas de

información geográfica, visualización geográfica, percepción remota y geomática. También ha trabajado en aplicaciones de HCI y sistemas contextuales.

El Dr. Sergio Zepeda-Hernández realizó su doctorado en Ingeniería Eléctrica en la especialidad de Informática del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (México) en 2009. Actualmente es profesor-investigador del Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Sus intereses de investigación incluyen la ingeniería web, recuperación de información, Interacción Humano Computadora, usabilidad y diseño de interfaces.