

# The Mass Media bias: Analysing and comparing the time series of polls and news articles during the 2016 USA presidential election.

Albanese F<sup>1,2,†</sup>, Balenzuela P<sup>1</sup>, and Semeshenko V<sup>3</sup>

<sup>1</sup>Departamento de Física, Universidad de Buenos Aires

<sup>2</sup>Departamento de Computación, Universidad de Buenos Aires

<sup>3</sup>Instituto Interdisciplinario de Economía Política, Universidad de Buenos Aires

<sup>†</sup>ffalbanese@gmail.com

May 2018

## Palabras Claves

Análisis de sentimiento; Detección de tópicos; Minería de textos; Dinámica de opinión.

## Resumen extendido

Día a día, las personas nos informamos consumiendo noticias de diferentes medios. Estudios empíricos demuestran que la dinámica de formación de opiniones durante períodos electorales está fuertemente influenciada por los medios de comunicación masivos [1]. Desde otro enfoque, las investigaciones teóricas realizadas en los últimos años muestran también cómo los medios pueden repercutir de forma directa en una sociedad [2]. En este trabajo se realizó un estudio teórico-empírico del rol político de los medios de comunicación utilizando técnicas de procesamiento del lenguaje natural y métodos estadísticos.

Para ello, se analizaron todas las frases provenientes de todas las noticias relacionadas con los candidatos procedentes de grandes medios de comunicación (The New York Times, Fox News y Breitbart) durante el período pre-electoral estadounidense del 2016 entre los candidatos Donald Trump y Hillary Clinton (28/07/2016 a 08/11/2016).

Se utilizaron algoritmos de detección de tópicos en textos (Latent Dirichlet Allocation [3] junto a Term frequency – Inverse document frequency [4]) y modelos profundos recursivos para la composición semántica aplicado sobre árboles

de sentimiento [5], para conocer si las menciones de los candidatos son de carácter positivo o negativo. De esta forma se buscó caracterizar de forma automática el flujo de noticias de cada medio y su relación con el apoyo a los candidatos. En particular, testear las hipótesis de que la propagada negativa de un medio sobre un candidato influye negativamente en dicha persona y positivamente en el rival.

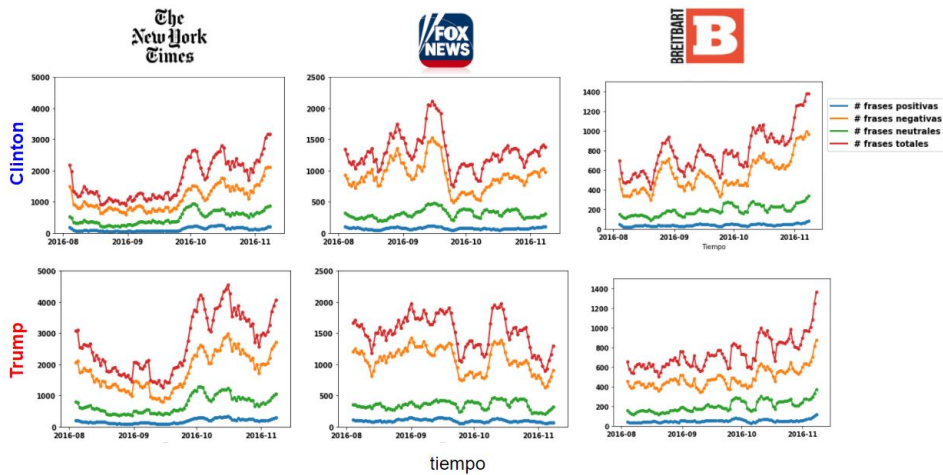


Figura 1: Curvas generadas a partir de la cantidad de menciones positivas (azul), negativas(amarillo), neutras(verde) y totales (rojo) de los candidatos para cada uno de los medios.

En la figura 1 se pueden observar las series de tiempo correspondientes a los análisis de sentimiento para ambos candidatos separados por medio de comunicación. Para generar las mismas, todas las frases de todas las noticias correspondientes al periodo pre electoral fueron tenidas en cuenta. En primer lugar, se puede observar una clara preponderancia de menciones negativas para todos los medios y ambos candidatos. En base a estudios complementarios, este efecto puede deberse a la naturaleza negativa de las noticias.

Al comparar dichas series de tiempo con los resultados de 263 encuestas nacionales realizadas por distintos organismos en función del tiempo [6] (Figura 2), se logro observar de forma estadísticamente significativa (p-valor menor a 0.05) una relación entre lo que los artículos de las noticias y el escenario político durante las elecciones presidenciales. En particular, teniendo en cuenta el signo del coeficiente de Pearson de la correlación, se puede confirmar el sesgo de los medios: Cuando el New York Times distribuye un artículo, Hillary Clinton sube en las encuestas, independientemente del sentimiento del texto y del candidato que se menciona. Por el contrario, cuando Fox News distribuye un artículo, Donald Trump sube en las encuestas.

Además, al realizar un análisis de la entropía e información mutua [7], se

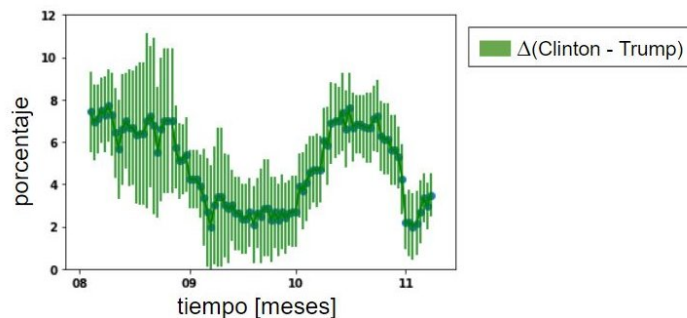


Figura 2: Diferencia entre las curvas de encuestas de Clinton y Trump.

concluyo que la positividad y la negatividad de las menciones es relevante de forma no lineal. En particular y como ejemplo para clarificar, se vio que tanto frases positivas de Clinton como negativas de Trump en los diarios correspondían con que dicha candidata subiera en las encuestas.

## Referencias

- [1] Alan S Gerber, Dean Karlan, and Daniel Bergan. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, 2009.
- [2] Sebastián Pinto, Pablo Balenzuela, and Claudio O Dorso. Setting the agenda: Different strategies of a mass media in a model of cultural dissemination. *Physica A: Statistical Mechanics and its Applications*, 458:378–390, 2016.
- [3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [4] Juan Ramos et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, volume 242, pages 133–142, 2003.
- [5] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.
- [6] Real clear politics. <https://realclearpolitics.com/>.
- [7] Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 88(17):174102, 2002.