

1 de Junio de 2017

La demoscopia en la era del Big Data

Utilizando los análisis de tendencias en internet para predecir resultados electorales. Tres casos prácticos.

Alejandro Souren

Tutor: Narcís Iglesias

Estudio: Grau en Comunicació Cultural

Facultat de Lletres

Universitat de Girona

Contenido

1. Introducción:	3
1.1 Aclaraciones previas	5
2. Estado de la cuestión y marco teórico:	8
2.1. Las implicaciones del análisis del Big Data	8
2.2 Referentes teóricos escasos.....	10
3. Metodología:	13
4. Objetivos e Hipótesis:	17
5. Análisis:	19
5.1 El caso británico: Referéndum del 23 de junio del 2016 sobre la situación del Reino Unido en la Unión Europea.	19
5.2 El caso estadounidense: Elecciones presidenciales de noviembre del 2016.....	28
5.3 El caso francés: elecciones presidenciales de abril-mayo 2017.....	37
5.3.1 La primera vuelta:	38
5.3.2 La segunda vuelta	47
6. Conclusiones:	53
6.1 Reflexiones finales.....	54
7. Bibliografía	57
8. Apéndice	59

1. Introducción:

El 9 de noviembre del 2016, en vísperas de la victoria electoral de Donald J. Trump, y cuando los primeros recuentos de votos apuntaban a un revés imprevisto, el *New York Times* publicaba un artículo titulado “How Did the Media — How Did We — Get This Wrong?” (Barbaro, 2016). El artículo en cuestión venía encabezado por una imagen de la redacción del periódico a las 3:30 de la madrugada; la conmoción de los periodistas del periódico más notable y respetado del planeta era visible, y alzaba la pregunta obligada a la que nadie se quería enfrentar: ¿Por qué, nosotros, la prensa, con todos los recursos de los que disponemos, no vimos llegar unos resultados tan abrumadores? Michael Barbaro, autor del artículo, se cuestiona en el mismo: “How did he pull off such a stunning victory? How did almost no one — not the pundits, not the pollsters, not us in the media — see it coming?” (Barbaro, 2016). El artículo del *New York Times* fue solamente el primero de muchos otros que se publicaron a raíz de la inesperada victoria electoral de Donald J. Trump. La prensa de medio mundo se levantó la mañana del 9 de noviembre preguntándose el porqué de un fracaso tan sonado; y cómo la demoscopia podría superar aquel golpe. No obstante, la sorpresa general causada por la victoria electoral del actual presidente de Estados Unidos no fue un caso aislado. Artículos similares se publicaron a lo largo del 2016, un año en el que la demoscopia tradicional no hacía más que incurrir en los mismos errores: dar una ventaja holgada a opciones no ganadoras. Lo mismo sucedió el 23 de junio del 2016, cuando la conmoción se repitió, esta vez en el Reino Unido. Casi el 52% de los participantes del referéndum sobre la situación del Reino Unido en la Unión Europea decidieron que abandonar la Unión era la mejor alternativa para su país. Los defensores del *remain* (quedarse en la UE) perdieron, para sorpresa de todas las encuestas nacionales e internacionales. Y los ejemplos se suceden. Algo parecido sucedió en Colombia el mismo año, donde el 4 de octubre el 50,2% de la población decidió que no apoyaba la paz entre el gobierno y las guerrillas de las FARC. Pero también aquí, en España, vivimos en el 2016 un episodio inesperado para las encuestas. La gran mayoría de medios tradicionales, salvo excepciones, preveían un *sorpasso*¹ de la formación política Podemos al Partido Socialista Obrero Español en las elecciones generales del 26 de junio. Para sorpresa de muchos, el *sorpasso* nunca se produjo y, además, Podemos sufrió un importante hundimiento en los resultados respecto a las previsiones demoscópicas anteriores a la celebración de elecciones. Al margen de los factores que provocaron todos estos inesperados desenlaces, el 2016 dejó claro un hecho que llevaba ya algunos años anunciándose: la demoscopia, tal y como la conocemos hoy, ha fracasado, y con ella la prensa que la secunda. Como apunta Agustín Giménez (2016) en un artículo sobre la cuestión para la ANN:

“Los tres sucesos electorales más importantes del 2016 (Brexit, No a la Paz y el triunfo de Trump) tuvieron factores en común que pueden señalar y ahondar en las fallas que los obsoletos

¹ Se entiende por *Sorpasso*, el fenómeno por el cual un país o partido político supera a otro en términos económicos o políticos.

modelos de encuestas tuvieron en cada una de estas contiendas: los resultados fueron opuestos a los que medios de comunicación masivos, encuestadores, consultores y periodistas sostuvieron hasta hora antes de cada sufragio; los ciudadanos se expresan en grupos cerrados o plataformas sociales donde los analistas tradicionales no tienen acceso; el anonimato a la hora de responder preguntas fue clave para que los analistas de big data sí comprendieran las tendencias mientras que los encuestadores seguían trabajando con paneles clásicos.”

El declive de la demoscopia coincide, además, con la emergencia de nuevas formas de comunicación, a saber, las redes sociales y las distintas plataformas de búsqueda en internet. El 2015 supuso un punto de inflexión para la historia de la red, pues se superó la barrera de los 3.000 millones de usuarios en todo el mundo (Sociedad de la Información en España, 2016, pág. 37). Si bien cada uno de estos usuarios actúa de una forma diferente en la red, todos ellos desprenden una serie de metadatos al navegar: ubicación, páginas visitadas, historial de navegación en la red, enlaces pinchados, etc. Todos estos datos son almacenados en los servidores de los distintos proveedores de red, y pueden ser rastreados, analizados y estudiados. El interés que puedan inducir estos datos de forma individual se reduce a los cuerpos de seguridad y los servicios de inteligencia durante una investigación criminal. No obstante, es el potencial de la totalidad de estos datos la que alza nuevas posibilidades intrigantes. Pero para entender con claridad el potencial de estos datos, necesitamos primero entender cuál es el porcentaje de la población que los genera.

Un 78,7 % de la población española, entre los 16 y los 74 años, utilizan internet de forma frecuente, lo que corresponde a más de 27,15 millones de habitantes (Sociedad de la Información en España, 2016). De estos, un 81% afirma utilizar las redes sociales de forma frecuente (Arribas & Montanera, 2016). El promedio dedicado a las redes sociales en España asciende a las tres horas diarias, y son cada vez más los usuarios que utilizan la red para buscar información y compartir opiniones con miembros de un mismo círculo social. Las redes sociales han desplazado a las reuniones destinadas a las opiniones políticas de antaño y las han convertido en una parte de la vida cotidiana.

Una de las posibilidades más intrigantes alzada por la emergencia de las redes sociales es su uso para reemplazar los métodos tradicionales de demoscopia. Si en el último año hemos visto como los métodos clásicos de encuesta han fallado en su labor de predecir resultados electorales, lo contrario ha sucedido con el análisis de macrodatos. Para referirme a este último usaré el término “demoscopia virtual” a lo largo de este trabajo.²

Sin embargo, antes de continuar me gustaría aclarar algunos hechos que podrían llevar a confusión. Para comprender con totalidad lo que significa el Big Data, es importante que aclaremos algunas asunciones erróneas.

² El término “demoscopia virtual” no existe como tal y no se ha estandarizado para referirse al análisis de macrodatos. No obstante, en aras de la simplicidad, me he tomado la licencia de utilizar este término de creación propia para referirme a los métodos para predecir elecciones que utilizan el análisis de Big Data.

1.1 Aclaraciones previas

Generalmente, cuando hablamos de Big Data, pensamos únicamente en los datos que desprendemos en las redes sociales: nuestra edad, nuestro nombre, nuestra dirección y quizá nuestro número de teléfono. Si bien todos estos datos suponen una parte importante de lo que llamamos Big Data, no conforman la totalidad de esta información. El Big Data se compone de distintos elementos procedentes de diferentes fuentes. Entre estas fuentes se encuentran nuestros teléfonos móviles, ordenadores, *tablets*, *wearables*, electrodomésticos conectados, coches con ordenadores a bordo y, en general, cualquier dispositivo doméstico como bombillas o cerraduras inteligentes. El Big Data, en definitiva, se genera a partir de todo dispositivo que tenga la capacidad de conectarse a Internet. Naturalmente, la parte más importante de estos datos es la que generan los usuarios en el momento de conectarse a la red, a saber, en redes sociales, bancos online o proveedores de correo electrónico. Debemos, por tanto, olvidar la noción de que el Big Data se conforma únicamente de publicaciones en redes sociales y mensajes de texto.

Todos estos datos se almacenan en servidores, normalmente privados, y pueden ser analizados posteriormente. Naturalmente, la cantidad de datos generada a diario impide que estos datos sean monitorizados en su totalidad en tiempo real; no obstante, es fácil observar tendencias. Para ello, compañías como Google han creado herramientas específicas, algunas de ellas de acceso público, para buscar y analizar tendencias.³ De hecho, los casos prácticos analizados en este trabajo hacen uso de estas herramientas.

Considero también importante aclarar que la tesis de este trabajo no se basa en la predicción electoral únicamente a través de datos recolectados de redes sociales como Twitter o Facebook, sino que plantea la tangible posibilidad de utilizar las herramientas de análisis de Big Data de acceso abierto para predecir resultados electorales con certeza y crear, de esta forma, una nueva forma de *demoscopia virtual* que pueda servir de herramienta al periodismo de los próximos años. El motivo por el que la predicción únicamente a través de redes sociales ha quedado desechada en este trabajo, lo explica con claridad el doctor Taha Yasseri del *Oxford Internet Institute* y autor del estudio “Can electoral popularity be predicted using socially generated Big Data?” En este mismo texto, el profesor Yasseri (2014, p.3) nos explica los principales motivos por los que confiar únicamente en las redes sociales puede ser una mala idea si lo que queremos es predecir resultados electorales con exactitud.

Como apunta Yasseri, una de las posibilidades más esperadas de la emergencia de las redes sociales fue, sin duda, su capacidad de complementar o incluso suplir los métodos tradicionales de demoscopia, ya que las redes sociales ofrecen considerables ventajas en comparación con las encuestas, sobre todo en términos de velocidad y recolección de

³ Me remito al servicio de acceso libre *Google Trends*, disponible en: <https://trends.google.es/trends/>

datos. Estas características hacían de las predicciones a través de redes sociales un sistema mucho más económico y rápido que la demoscopia tradicional. En un principio, el uso de estos nuevos métodos de predicción electoral fue solamente eso, un método para predecir elecciones. No obstante, el potencial de la predicción a través de redes sociales empezó a ofrecer nuevos campos donde ponerlos en práctica. En poco tiempo, la predicción a través de las redes sociales se aplicó a terremotos o incluso atascos de tráfico⁴.

Existen dos métodos para extraer una predicción de redes sociales como Twitter. El primer método consiste en contar los tuits o publicaciones que mencionen al candidato en cuestión. Y el segundo, más frecuente, utiliza métodos elaborados de interpretación de las publicaciones que se realizan en una red social⁵. Sin embargo, a pesar del entusiasmo inicial y en contraste con la predicción de terremotos o atascos, la mayoría de estudios que han intentado usar redes sociales como Twitter para predecir resultados electorales han fracasado. Muchos investigadores acabaron llegando a la conclusión de que había poca correlación entre la popularidad del candidato en la red y los resultados definitivos y que la mayor parte de los aciertos podrían haberse dado por casualidad.

El sesgo selectivo en la predicción basada en redes sociales como Twitter parece claro: no todo el mundo usa redes sociales como Twitter, y los que la usan no están aleatoriamente distribuidos a través de la población. Hay grupos sociales mucho más activos que otros en la red, y no por ello representan parte de la mayoría electoral. A nivel nacional, cabría destacar que este fue uno de los principales motivos por los que el esperado *sorpas* del partido político Podemos nunca se produjera. Los investigadores que utilizaron las redes sociales para analizar tendencias de comportamiento en la red, delegaron gran parte de sus fuentes en las redes sociales como Twitter, sin tener en cuenta que la mayoría de los usuarios de Twitter son precisamente jóvenes y, por ende, más propensos a votar a partidos de izquierdas como Podemos. No obstante, los jóvenes no conforman la mayoría del censo español.

Como apunta el profesor Yasseri, este sesgo parece ser endémico en los estudios que utilizan solamente redes sociales para hacer predicciones. Es frecuente que los resultados de los estudios en cuestión presenten una infrarrepresentación de partidos minoritarios y prefieran a los partidos considerados progresistas o de izquierdas en detrimento de los conservadores (Yasseri, 2014, pág. 3).

Se han presentado, en los últimos años, una variedad de plausibles explicaciones a estos hechos. El más obvio parece ser el de la “autocensura”; en efecto, es en última instancia

⁴ Véase: Sakaki T, Okazaki M, Matsuo Y., “Earthquake shakes Twitter users: real-time event detection by social sensors.” In: *Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, WWW '10, pp. 851–860, (2010).*

⁵ Véase: O’Connor B, Balasubramanyan R, Routledge BR, et al., “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, Washington, DC, 23-26 May, (2010).*

el usuario el que decide, o no, publicar una opinión en la red social. De esta manera, no son pocos los usuarios que moderan sus opiniones en función del tipo de relaciones sociales que tengan con el resto de usuarios de la red. Asimismo, algunos investigadores han observado la dificultad de analizar correctamente el sentido de un tuit, tanto por la cantidad de información limitada que permite (140 caracteres) y por el muy frecuente uso de la ironía o el sarcasmo que los usuarios dan en las redes sociales.

Por último, con la prominencia cada vez mayor de las redes sociales en el panorama mediático, son numerosos los candidatos políticos que han empezado a intervenir de forma activa en las redes. Cabe destacar que este último hecho les permite alterar o manipular las opiniones que los usuarios de una red social publican; a saber, con una campaña social fuerte que cree tendencias (*Trending Topics*) en una red social. Esto último también supone un sesgo importante para la predicción electoral basada solamente en redes sociales.

2. Estado de la cuestión y marco teórico:

2.1. Las implicaciones del análisis del Big Data

El análisis del Big Data es un campo de estudio todavía muy reciente. De hecho, el propio término Big Data no dispone de una traducción estandarizada al castellano; se manejan entre posibles traducciones *macrodatos* o *inteligencia de datos*, pero la voz inglesa se ha impuesto en casi todo el mundo académico. Ello se debe, sobre todo, porque las primeras investigaciones realizadas sobre este concepto, así como su aplicación a distintas disciplinas, se han llevado a cabo en el mundo anglosajón, especialmente en Estados Unidos. Sin embargo, si bien los estudios realizados sobre estos datos, que los usuarios desprenden mientras navegan por la red, todavía son muy reducidos, sus aplicaciones parecen ser interminables. Desde la publicidad, pasando por la política y hasta el periodismo, el estudio del Big Data ofrece innumerables soluciones a problemas actuales. No son pocas las agencias de publicidad que utilizan el análisis de *macrodatos* para seleccionar un *target* (público potencial) específico, y de esta forma abaratar los costes que supone la publicidad al por mayor. También la política ha encontrado una solución a sus problemas en el análisis del Big Data. En las elecciones presidenciales del año 2008 el gabinete de campaña del por entonces candidato demócrata, Barack Obama, contrató a uno de los mayores expertos en análisis de Big Data, el paquistaní Rayid Ghani, *Chief Data Scientist* en la Universidad de Chicago (Peña, 2016). La popularidad de Obama en las redes sociales fue un factor determinante en su victoria electoral y mandó un claro mensaje a la política del siglo XXI; el camino a seguir se encontraba en el análisis del Big Data.

El periodismo, por otro lado, también ha encontrado en el análisis de macrodatos una nueva forma de trabajar. El análisis del Big Data permite a los periodistas observar la evolución de sus propias noticias en tiempo real, y medir el impacto que tienen en determinados tipos de usuarios. Además, ofrece la posibilidad de adaptar las publicaciones al interés general de la audiencia dependiendo del tipo de usuario al que se quiera atraer. Incluso permite saber dónde y en qué momento se está desarrollando una historia válida para ser cubierta.

Sin embargo, el periodismo también ha encontrado en el Big Data una amenaza a su existencia. Son cada vez más los algoritmos que permiten la redacción automatizada de noticias, algo que amenaza, sobre todo, a los periodistas de agencias de información. Si bien la faceta artística de los algoritmos todavía no está lo suficientemente desarrollada como para substituir la faena de un periodista, sí han demostrado ser capaces de redactar

textos con un increíble parecido a los textos redactados por humanos, e incluso algún poema⁶ (Sánchez, 2016).

No obstante, sobre la implicación del Big Data y los algoritmos en la labor del periodista se han escrito ya algunos ensayos académicos. Considero especialmente relevante el texto Seth C. Lewis, “Journalism In an Era Of Big Data”, donde el propio autor deja clara la importancia que pueden tener los algoritmos en la misma naturaleza de la redacción periodística. Citando a Matt Carlson nos dice:

“[...] Matt Carlson explains what begins to happen as “the role of big data in journalism shifts from reporting tool to the generation of news content” in the form of what he calls “automated journalism.” The term refers to “algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming.” Among the data-oriented practices emerging in journalism, he says, “none appear to be as potentially disruptive as automated journalism,” insofar as it calls up concerns about the future of journalistic labor, news compositional forms, and the very foundation of journalistic authority [...]” (Lewis, 2015, pág. 326)

Sobre la percepción pública de estos mismos textos periodísticos redactados por algoritmos considero relevante el estudio realizado por varios autores coreanos titulado: “Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists”. Publicado en febrero de 2017. Como apuntan sus autores, en relación a lo citado anteriormente: “*In this new era of automated journalism, to activate discussion on automated news creation and its implications for journalists, journalism, news industry, and audience, it is necessary to explore the public's and journalists' perceptions and evaluation of news written by algorithms.*” (Jung, Song, Kim, Im, & Oh, 2016, pág. 292)

Estas noticias, que inevitablemente serán una realidad en un futuro no muy lejano, también despiertan debates de tipo ético. Sobre este asunto los profesores Konstantin Nicholas Dörr y Katharina Hollnbuchner de la Universidad de Zurich han publicado un interesante artículo titulado “Ethical Challenges of Algorithmic Journalism” (Dörr & Hollnbuchner, 2016)⁷

Los tres casos de estudio presentados resultan esenciales para comprender hacia qué dirección se dirige el periodismo de los próximos años. Existen, al menos entre ciertos periodistas, el miedo de que algún día sus puestos de trabajos sean sustituidos por máquinas. No es descabellado plantear esta realidad, pues ya se han visto precedentes como el de *Los Angeles Times* (BBC, 2014). Pero el entusiasmo o miedo que haya podido

⁶ Me remito a sistemas de creación “artística” automatizada como WASP o ASPERA, en el caso de la lengua castellana. El desarrollador de estos sistemas, Pablo Gervás, redactó un informe sobre el sistema y su funcionamiento en el 2016. Se encuentra disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.3084&rep=rep1&type=pdf>

⁷ Disponible en: <http://www.tandfonline.com/doi/full/10.1080/21670811.2016.1167612>

generar esta capacidad creativa de la tecnología, contrasta con el desinterés mostrado por el potencial predictivo del Big Data. Si bien es cierto que en los últimos años ha aparecido lo que se denomina el “periodismo de datos”, este sigue siendo un ámbito muy concreto del periodismo de investigación del que se han publicado pocos estudios. Sobre la capacidad predictiva de los macrodatos versa este trabajo.

2.2 Referentes teóricos escasos

Si bien es cierto que la aplicación del Big Data en distintos ámbitos de la sociedad es cada vez más una realidad, no podemos decir lo mismo de su estudio por parte del mundo académico. El ámbito académico anglosajón, como hemos visto, ha empezado en los últimos años a interesarse por las posibles implicaciones del Big Data y ha publicado algunos estudios relevantes sobre la cuestión. Sin embargo, estos estudios versan sobre aplicaciones reales y tangibles de los macrodatos⁸, y son pocos los autores que se aventuran a estudiar el potencial predictivo del Big Data.

La única aplicación posible del Big Data al mundo del periodismo que parece interesar a los especialistas, es la de su capacidad para redactar una noticia periodística. Sin embargo, el análisis de macrodatos también abre la puerta a su interpretación como una base predictiva fiable. Sobre esta última aplicación existen pocos trabajos académicos a los que recurrir, y los que existen salen del mundo anglosajón. Actualmente, recurrir al Big Data para obtener predicciones parece más el argumento de una película de ciencia ficción como *Minority Report* que una posibilidad real y ponderable. No obstante, existen especialistas que han intentado estudiar este método aplicado, por ejemplo, a elecciones políticas.

Los referentes directos de este trabajo se pueden resumir en dos estudios sobre la cuestión. En primer lugar, el ejercicio presentado en este trabajo se inspira, en gran medida, en el trabajo del profesor Taha Yasseri del *Oxford Internet Institute* titulado “Can electoral popularity be predicted using Big Data and network analysis?”. El estudio, publicado en el 2014, analiza la posibilidad de predecir resultados electorales con cierta exactitud usando como herramientas el análisis de tendencias en el buscador de internet Google y cruzándolos con el número de visitas de las páginas de Wikipedia de los partidos y/o candidatos. Para hacerlo, el profesor Yasseri se basó en tres casos prácticos: las elecciones del 2013 en el Reino Unido, Alemania e Irán. Estos tres países disponen de sistemas políticos distintos y por tanto el ejercicio ofrecería una visión global de la predicción electoral a través de la red en diferentes contextos. Los resultados del estudio fueron alentadores: de los tres casos, el mejor resultado se obtuvo en las elecciones alemanas, donde las tendencias de Google predijeron, casi con exactitud, la primera y segunda posición de las elecciones. Ocurrió lo mismo en el caso iraní, donde ambas plataformas predijeron la victoria de Rouhani; sin embargo, fracasaron en predecir la segunda posición. Por último, el caso inglés resultó ser más decepcionante. Ambas plataformas

⁸ Como los mencionados en el apartado anterior.

sobreestimaron al partido Liberal demócrata y no consiguieron predecir los resultados con un mínimo de exactitud.

El experimento del profesor Yasseri, a pesar de no haber sido capaz de demostrar la hipótesis al completo, ofreció dos importantes conclusiones:

1. Durante periodos electorales crece enormemente el tráfico en las redes. Tanto en redes sociales como en páginas informativas como Wikipedia (Yasseri, 2014, pág. 7).
2. La predicción por vía de analizar tendencias en internet funciona mucho mejor en el caso de elecciones donde se potencia la figura individual del candidato por encima del partido (Yasseri, 2014, pág. 7).

Esta última conclusión resulta de vital importancia para la naturaleza del presente trabajo, que puede verse como una continuación de los estudio del profesor Taha Yasseri.

Por otro lado, pero en menor medida, han sido de gran ayuda los trabajos de Catherine Lui, Takis Metaxas y Eni Mustafari y su estudio para el Wellesley College titulado “On the predictability of the U.S. elections through search volume activity” (Lui, Metaxas, & Mustafaraj, 2011). En este trabajo se intenta predecir los resultados de las elecciones al Senado de Estados Unidos en los años 2008 y 2010. A pesar de las posibilidades que ya en el 2011 había demostrado la predicción a través de internet, los resultados del estudio en cuestión fueron pobres. El análisis de tendencias solamente consiguió predecir el 43% de los candidatos seleccionados en el 2008 y el 40% en el 2010, algo que podría explicarse casi a través de la estadística (Lui, Metaxas, & Mustafaraj, 2011, pág. 7). El estudio fracasó en sus objetivos⁹, pero dejó espacio a algunas conclusiones interesantes que han servido para la realización del presente trabajo.

1. Las predicciones más acertadas se realizaron con los candidatos que no utilizaban redes sociales. Eso seguramente se deba a que creaban menos contenidos polémicos capaces de adulterar los resultados.
2. Google Trends demuestra ser una herramienta compleja para predecir procesos electorales, pues es imposible saber la opinión de la persona que introduce un término de búsqueda concreto. Y es casi imposible comprender las circunstancias por las cuales se busca el nombre de un determinado candidato.

Estas conclusiones han sido importantes porque me han permitido enmendar algunos de los errores incurridos en anteriores estudios en este trabajo. En primer lugar, y como figura en el apartado de metodología de este trabajo, las búsquedas en Google Trends han sido filtradas y cruzadas con términos más precisos que solamente el nombre del candidato (más obre esto en el apartado Metodología de este trabajo). De esta forma, no

⁹ Los autores afirman: “*So, overall, G-trends was not a good predictor of both the 2008 and 2010 US congressional elections. Still, this performance begs some explanation*” (Lui, Metaxas, & Mustafaraj, 2011, pág. 8)

solamente se analiza la tendencia de búsqueda del nombre de un candidato, sino también las búsquedas relacionadas, para determinar si la percepción es positiva o negativa.

A pesar de las conclusiones delusorias a las que llegaron los estudios que se han utilizado como referencia en este trabajo, he decidido continuar adelante con el uso de Google Trends para realizar mis análisis. Esta decisión responde, principalmente, a dos criterios.

En primer lugar, hemos visto como en los últimos años se ha incrementado exponencialmente el número de usuarios que hay en la red, y en consecuencia la información que en ella se publica. Además, posiblemente, gran parte de los internautas jóvenes que en el 2011 y 2014 todavía no podían votar, hoy sí lo hacen. Es un hecho que a medida que la población joven crezca, una mayor parte del censo electoral tendrá presencia en la red.

Por otro lado, tanto el estudio del doctor Taha Yasseri como el de Catherine Lui se han realizado sobre procesos electorales abstractos; es decir, elecciones en las que las opciones disponibles admitían diversas variables. El propio profesor Taha Yasseri reconoce que la aplicación de los análisis de tendencias en internet es mucho más efectiva en sistemas electorales donde se potencian opciones binarias o muy concretas (Yasseri, 2014, pág. 7).

3. Metodología:

Para el análisis de los tres casos prácticos que aparecen en este trabajo he utilizado dos herramientas de acceso abierto disponibles en la web: Google Trends¹⁰ y una interfaz gráfica del contabilizador de visitas de páginas de Wikipedia¹¹. La elección de ambas páginas responde a la popularidad de la que gozan ambos servicios, siendo, además, las referencias para buscar información en la red.

Google se ha convertido en pocos años en el buscador hegemónico de la red. Es, a día de hoy, la forma paradigmática de búsqueda en internet y ha desarrollado algoritmos únicos para acceder a la información. El dominio de Google, de hecho, tiene un impacto sustancial en la manera en la que las personas buscamos información en internet. Si bien es cierto que algunas empresas han intentado presentar alternativas viables al buscador de Google, ninguna ha conseguido igualar sus resultados. Este hecho ha provocado que en los últimos años otros buscadores de internet hayan intentado replicar los algoritmos de Google (Lui, Metaxas, & Mustafaraj, 2011, pág. 3). Además, el hecho de ser el buscador más utilizado en la red nos ofrece unos resultados más precisos, al englobar una mayor parte del censo potencial.

Para realizar el ejercicio, se ha hecho uso de la herramienta Google Trends, creada en el 2006. La herramienta permite que los usuarios puedan observar y analizar tendencias de búsqueda en todos los idiomas, servicios y ámbitos geográficos en los que se encuentra disponible el buscador. El funcionamiento del servicio es el siguiente. El buscador analiza una porción de las búsquedas en la red para computar cuantas se han realizado con los términos introducidos por el usuario en relación al total de las búsquedas realizadas en un lapso de tiempo determinado. Es decir, cuando el usuario introduce un término de búsqueda en Google Trends, el servicio muestra una gráfica en la que se muestra el incremento o el descenso de las búsquedas de aquel término en un periodo de tiempo especificado por el usuario. Es importante destacar que, como los datos de búsquedas en bruto son información sensible, Google no ofrece números exactos. En su lugar, el servicio devuelve una cantidad ponderada¹² al usuario, que la herramienta compara con las búsquedas normales del mismo término. Por tanto, el uso de Google Trends no nos ofrece datos de búsqueda en términos absolutos, pero sí nos ofrece una tendencia clara, indicando el porcentaje de aumento o descenso en búsquedas.

Un estudio realizado en el 2009 por Choi Hyunyoung y Hal Varian muestran cómo es posible utilizar dicho servicio para predecir la demanda de bienes como coches o viviendas. El estudio parte de la base de que el número de búsquedas en internet en un determinado ámbito geográfico es un claro indicador de que existe un interés por adquirir esos bienes (Choi & Varian, 2009). El estudio ofreció resultados fiables que coincidían

¹⁰ Disponible en: <https://trends.google.es/trends/>

¹¹ Disponible en: <https://tools.wmflabs.org/pageviews/?project=en.wikipedia.org&platform=all-access&agent=user&range=latest-20&pages=CatIDog>

¹² Los datos se expresan del 1 al 100. Siendo 1 el mínimo y 100 el máximo.

con los obtenidos por otros métodos de análisis tradicionales. Además, el uso de herramientas como Google Trends demostró ser mucho más rápido, económico y accesible. Por otro lado, en la misma línea, el propio Google ha puesto en marcha un servicio llamado “Google Flu Trends” desde donde se analizan las tendencias de búsqueda de ciertos términos para predecir épocas de gripe¹³.

La segunda parte de los ejercicios prácticos presentados en este trabajo se han realizado con el contador de visitas de una página en Wikipedia. El sistema funciona de forma similar a Google, pero este sí ofrece datos de búsqueda en datos absolutos. Dicho de otro modo, Wikipedia nos ofrece números exactos de cuantas veces se ha visto una determinada página. Sin embargo, la forma en la que Wikipedia ofrece sus datos es inefable. El servicio oficial de Wikipedia descarga todos sus metadatos en una misma página; y navegar a través de ellos es una faena titánica. Por ello, y en aras de la simplicidad, han surgido numerosos servicios que se encargan de indexar esta información y presentarla en forma de gráfico. Para este trabajo se ha utilizado el servicio “Análisis de Visitas” desarrollado por tres usuarios anónimos de Wikipedia. Este servicio, sin embargo, ofrece solamente información detallada a partir de Julio del 2015, momento de creación de la herramienta. Este problema no ha afectado a los resultados de este trabajo pues ninguna fuente analizada es anterior a esa fecha.

Ambos servicios, tanto el de Google como el de Wikipedia, permiten a los usuarios descargar los datos requeridos en archivos “.csv” que después puede ser tratados con programas de ofimática como Excel.

Una de las razones por las que el número de visitas a la página de Wikipedia de un candidato resulta tan esclarecedor, es porque implican un interés por el candidato en cuestión. En este punto, las tendencias en Wikipedia resultan mucho más fiables que las tendencias de búsqueda genéricas en buscadores como Google. El motivo es claro: una persona que accede a la página de Wikipedia de una persona o tema, por definición, busca información sobre aquello que ha buscado. No ocurre lo mismo con Google, donde una búsqueda puede corresponder tanto al interés que suscita el tema, la búsqueda de una noticia en concreto o simplemente, una consulta gramatical (Gonzalo, 2013). Además, el hecho de que un individuo busque información sobre un determinado candidato en la Wikipedia durante un proceso electoral, implica que desea conocer más sobre este candidato, o, como poco, descubrir algo nuevo. Naturalmente no podemos tomar el interés de todos los usuarios que buscan información de un candidato como claros indicadores de voto, ya que muchos otros factores entran en juego a la hora de escoger. Aun así, no sería descabellado plantear que una gran parte de los usuarios que acceden a la página de Wikipedia de un candidato se planteen votar por él, o, al menos, obtener más información.

En el caso de Google Trends, las extrapolaciones son un tanto más complicadas. Si bien

¹³ El proyecto, descontinuado desde hace poco, está disponible en:
<https://www.google.org/flutrends/about/>

es cierto que una búsqueda única sobre un candidato puede indicar interés por el mismo, esta premisa es solamente una hipótesis. Los resultados que obtenemos cuando introducimos el nombre de un candidato electoral en Google son inabarcables, y Google no permite hacer un seguimiento de los resultados más visitados. Por tanto, solamente podemos suponer las páginas más visitadas por los usuarios. Esto es así porque por cada usuario que busque un nombre en Google para obtener información en una página como Wikipedia, habrá otro que lo busque para conocer las últimas noticias en un periódico digital. Es, por tanto, que un millón de búsquedas sobre un candidato no pueden servir de indicador electoral.

Sin embargo, un millón de búsquedas en internet sobre un candidato sí nos indicarían un hecho muy relevante para este trabajo: el aumento desmesurado del tráfico en internet durante periodos electorales. Algo que soportaría la tesis de que internet tendrá un papel clave durante los procesos electorales de los próximos años. Además, un mayor número de búsquedas en internet también demostrarían un aumento del interés por un candidato en particular; independientemente de la opinión que cada usuario tenga sobre el candidato en cuestión. Esto es algo que intentaré demostrar a lo largo de este trabajo. Para hacerlo cruzaré los resultados de Google Trends con los datos de visitas de Wikipedia, que, al ser más específicos, servirán de filtro. Además, en vista de que los resultados de Google Trends pueden ofrecer distintas lecturas, he optado por realizar dos análisis separados. Por un lado, como ya he explicado, he observado las tendencias de búsqueda de los candidatos en cuestión; esto es, el número de búsquedas en Google asociadas a su nombre. Esto es lo que Google denomina un “término de búsqueda”, es decir, un nombre concreto que redirige a una persona particular. De esta forma, un término concreto sería Donald J. Trump, 45º presidente de los Estados Unidos de América. Por otro lado, he comparado estos términos concretos con lo que Google denomina “temas de búsqueda”, es decir, palabras sueltas que no redirigen a nada en particular; para nuestro ejemplo, “*Vote Trump*” sería un tema de búsqueda.

El motivo por el que creo que es conveniente comparar ambas tendencias, es porque un tema de búsqueda como “*Vote Trump*” nos indicarían que la persona que lo introduce en el buscador está mucho más dispuesta y/o interesada a votar por el candidato que aquellos que solamente introducen su nombre. “Vote + candidato u opción”, por tanto, nos sirve de filtro para definir la cantidad de personas que buscan razones por las que votar a un candidato en particular. Añadir esta variable al análisis surge a raíz de las conclusiones del estudio de Catherine Lui, que afirma:

“A variable that may have affected G-trends effectiveness as a tool for predicting political elections is the sentiment of a user's query. It is difficult, though not impossible, to determine the circumstances behind a user's search of the profile of a certain candidate to make a guess about that candidate's public image and why a user might be interested in the candidate.” (Lui, Metaxas, & Mustafaraj, 2011, pág. 9)

Es con la intención de enmendar estos errores que he decidido incluir esta segunda variable en mi análisis de Google Trends. Es, por tanto, la primera vez que se realiza esta

operación para filtrar las búsquedas en un estudio de estas características.

Cabe destacar también que los análisis del caso británico y estadounidense se realizaron a posteriori, es decir, no hubo un seguimiento en tiempo real de los datos aquí analizados. El caso francés, en tanto que coincidió con la realización de este trabajo, ha sido monitorizado a tiempo real, con una frecuencia de dos veces por semana. Así lo reflejan también las gráficas presentadas en este trabajo. No obstante, en aras de la simplicidad, he decidido ordenar los tres casos prácticos de este trabajo por orden cronológico.

Por último, huelga decir que tanto los análisis de visitas de las páginas de Wikipedia como los análisis en Google Trends se han acotado a un ámbito geográfico concreto, esto es, el ámbito geográfico en el que se realizan las elecciones y en las que, por tanto, existe el mayor número de potenciales votantes. Y en un periodo de tiempo muy acotado también. A saber, a partir de las tres semanas anteriores a las elecciones, cuando la demoscopia es más activa.

4. Objetivos e Hipótesis:

El presente trabajo parte de la hipótesis de que es posible la predicción electoral por vía del análisis de Big Data con herramientas de acceso público. Para hacerlo me baso en tres casos prácticos alrededor de los que se articula este estudio: el referéndum sobre la situación del Reino Unido en la Unión Europea o Brexit (junio 2016); la victoria electoral de Donald J. Trump en las elecciones estadounidenses (noviembre 2016); y, por último, las elecciones generales francesas de abril del 2017. Los tres hechos anteriores tienen un componente común: una fuerte presencia en la red antes, durante y después de los procesos electorales.

El motivo por el que he seleccionado los tres acontecimientos en cuestión es porque todos ellos han tenido mucho eco mediático y han girado en torno a figuras fuertemente individualizadas. Este hecho ha sido imperativo para el correcto desempeño de este experimento, y parte de las conclusiones de los estudios realizados por el profesor Taha Yasseri de la universidad de Oxford anteriormente mencionados¹⁴. Además, los dos primeros casos analizados, el referéndum sobre el “Brexit” y la elección de Donald J. Trump, fueron acontecimientos que la demoscopia tradicional no fue capaz de predecir. El caso francés, sin embargo, sí fue vaticinado por las encuestas tradicionales y por tanto presenta un contrapunto importante que enriquece la tesis de este trabajo y ha sido seleccionado, entre otros motivos, para evitar el sesgo selectivo.

Tanto en el sistema político estadounidense como en el francés se refuerzan especialmente la figura del candidato individual por encima de sus partidos. Esto es de suma importancia cuando pretendemos utilizar las tendencias en internet como sistemas de predicción, pues los nombres de los candidatos nos tienen que servir de filtro. Esto último es importante porque una de las mayores dificultades para realizar este trabajo ha sido el de filtrar el contenido que se crea en la red para lograr una predicción mínimamente acertada.

Finalmente, la elección del caso británico tiene el mismo origen que los dos anteriores. La campaña por el referéndum sobre la situación de la Unión Europea en el Reino Unido estableció, desde sus inicios, dos opciones binarias muy claras: “*vote leave*” o “*vote remain*” y eso ha facilitado enormemente el filtrado de información.

Los objetivos principales de este trabajo se pueden resumir en cuatro, siendo estos los siguientes:

1. Demostrar que es posible predecir con cierta exactitud los procesos electorales en los que se potencia la figura individual del candidato a través del análisis de tendencias en internet. Para hacerlo presento tres casos prácticos.
2. Demostrar que estas predicciones y las herramientas utilizadas para realizarlas

¹⁴ El profesor Yasseri (2014, 7) sostiene: “*Overall, this may mean that predictions based on social data may perform better in political systems which encourage a focus on individuals.*”

pueden servir al periodista a ofrecer una información más completa del proceso electoral al lector o espectador.

3. Demostrar que estas herramientas tendrán una importancia imperiosa en la demoscopia de los próximos años.
4. Demostrar que si existe un aumento desmesurado del tráfico en internet durante periodos electorales esto hará de la red un elemento esencial para entender resultados electorales en los próximos años.

En definitiva, a través del análisis de tres casos prácticos intentaré demostrar que existen, a día de hoy, herramientas de acceso público que pueden pronosticar los resultados de ciertos procesos electorales de forma mínimamente acertada.

5. Análisis:

5.1 El caso británico: Referéndum del 23 de junio del 2016 sobre la situación del Reino Unido en la Unión Europea.

La relación entre el Reino Unido y la Unión Europea ha sido complicada desde sus inicios. Las islas británicas se adhirieron a la Comunidad Económica Europea el 1 de enero del 1973, bajo el gobierno conservador de Edward Heath. Esta decisión fue duramente criticada, entre otros, por el partido laborista, que se presentó a las elecciones de octubre del 1974 para renegociar los términos del tratado y celebrar un referéndum sobre su permanencia. En el 1975 se celebró finalmente un referéndum en el que se decidió solicitar la adhesión del Reino Unido al “mercado común europeo” o CEE. Las tensiones con Francia fueron constantes, sobre todo durante las administraciones de Charles De Gaulle. En el 1985, sin embargo, el Reino Unido decide no entrar en el espacio Schengen, pues prefiere mantener la soberanía de sus propias fronteras. Además de sus fronteras, el Reino Unido se niega a ceder su moneda, la libra esterlina, que seguirá siendo utilizada en el país hasta el día de hoy.

En el 2016, tras varias tensiones, sobre todo en materia de políticas sociales, y con el acecho del partido nacionalista UKIP, el entonces primer ministro, David Cameron, promete la celebración de un referéndum sobre la permanencia del Reino Unido en la unión. Tras una nueva victoria electoral, Cameron convoca un Referéndum para el día 23 de junio del 2016. Los resultados fueron ajustados pero claros, un 51,9% de la población decidió que querían abandonar la Unión Europea, frente al 48,1 que prefería quedarse (El País, 2016). Sin embargo, casi todas las encuestas oficiales publicadas vaticinaron una ajustada victoria de la opción “*remain*”, permanecer en la unión europea.

La campaña por el así denominado “Brexit” fue tensa desde sus inicios. Muchos culparon al por entonces primer ministro David Cameron de haber cedido a las presiones de la ultraderecha UKIP al convocar un referéndum; e, incluso, la campaña llegó a mancharse de sangre¹⁵. Sin embargo, a nivel informativo la campaña por el “Brexit” fue muy relevante. El referéndum sobre la permanencia en la Unión Europea fue especial por varios motivos, pero sobre todo porque estableció desde sus inicios dos opciones binarias muy claras: “*vote leave*” (abandonar la UE) o “*vote remain*” (permanecer en la UE). Este hecho, que puede parecer irrelevante en un principio, es de suma importancia cuando hablamos del análisis de datos en internet. Como apunta el profesor Yasseri, los sistemas de predicción electoral a través de tendencias en internet funcionan mejor en campañas donde las opciones están polarizadas y se resalta su carácter individual (Yasseri, 2014, pág. 7).

¹⁵ El 16 de junio del 2016 fue asesinada la diputada laborista Jo Cox, por un partidario de abandonar la Unión Europea radicalizado (Tubella, 2016).

Ambas campañas fueron muy parecidas durante todo el proceso. Entre los defensores de la opción “*Remain*” se encontraba el por entonces primer ministro David Cameron, y, en contra, el mayor beneficiado por la opción “*Leave*”, Nigel Farage, líder del partido nacionalista UKIP.

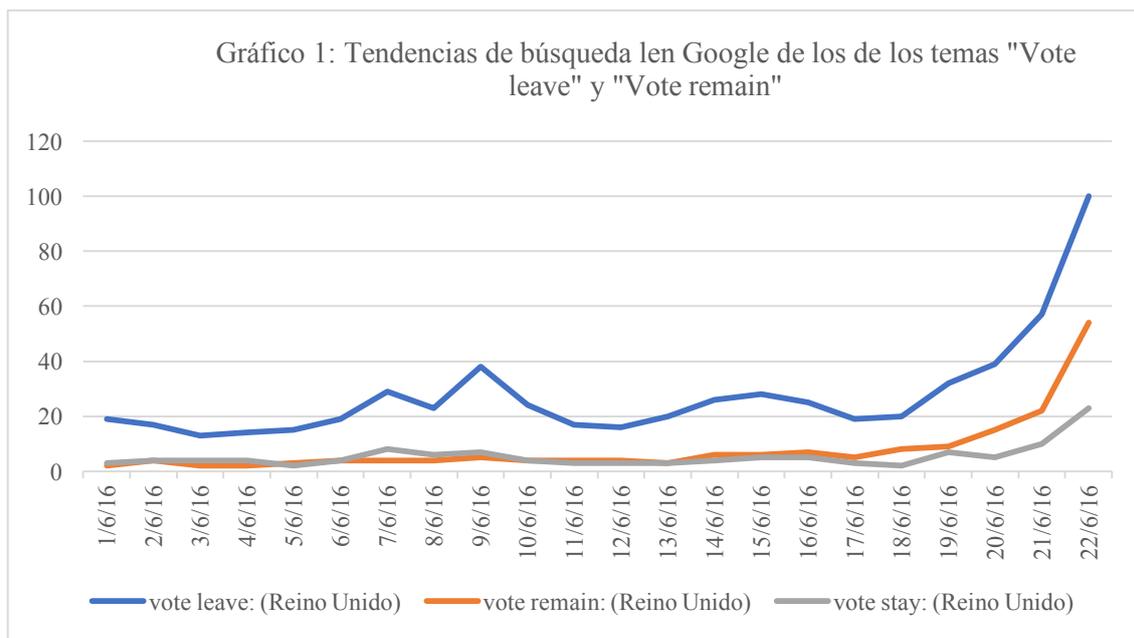
Debido a las enormes implicaciones que tendría un hipotético abandono de la Unión Europea por parte del Reino Unido, las predicciones demoscópicas fueron casi diarias. A pesar de la incertidumbre, la gran mayoría de encuestas vaticinaban una muy ajustada victoria de la opción “*Remain*” (permanecer); tal y como reflejan los datos siguientes del periódico *Financial Times*¹⁶:



Fuente: *Financial Times*

Sin embargo, si observamos los datos de tendencias en internet de Google Trends los resultados son distintos¹⁷ (gráfico 1).

¹⁶ Estos datos representan la media de todas las encuestas sobre el Referéndum que se realizaron. La zona gris central representa el porcentaje de la población indecisa y corresponde a un 6%. Fuente original: <https://ig.ft.com/sites/brexit-polling/>

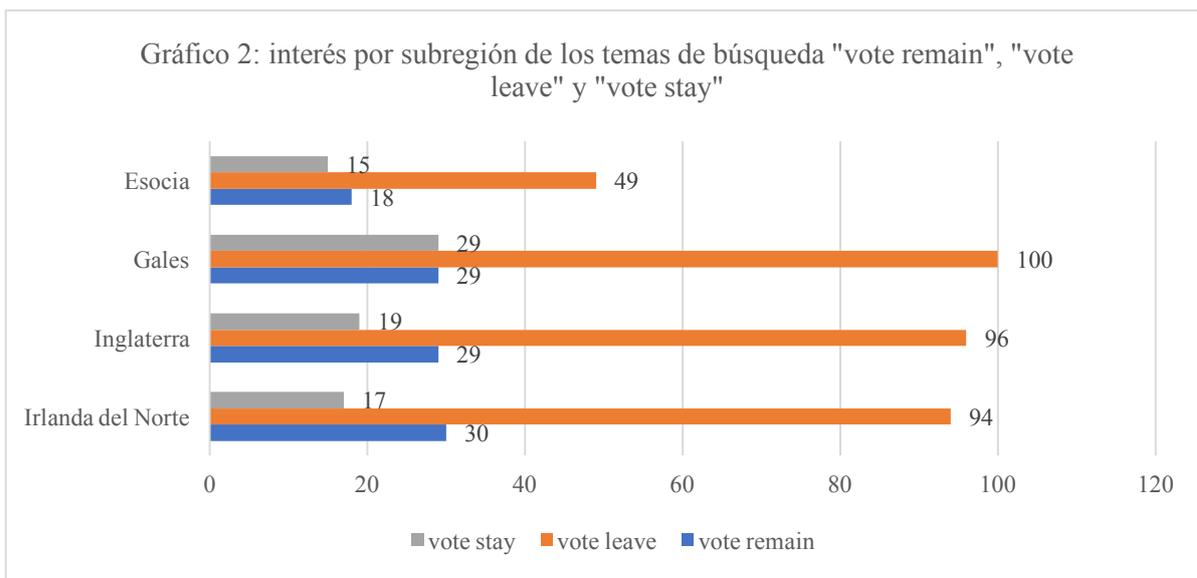


Fuente: Elaboración propia con los datos de Google Trends

Vemos en el gráfico superior como la opción de “*vote leave*” (abandonar) predomina durante las tres semanas anteriores al referéndum. El único descenso importante que se produce durante los días anteriores a la consulta es a partir del 16 de junio, seguramente a raíz del asesinato de la diputada laborista Jo Cox. Además, en el gráfico superior he incluido la tendencia de búsqueda de la opción “*stay*”, que, a pesar de no ser oficial, podría haber sido utilizada por algunos potenciales votantes. En ningún momento la suma de los valores de las dos opciones favorables a permanecer (*remain* + *stay*) llegan a superar el número de búsquedas originado por “*vote leave*”. No obstante, no incluir ambas opciones habría sesgado los resultados.

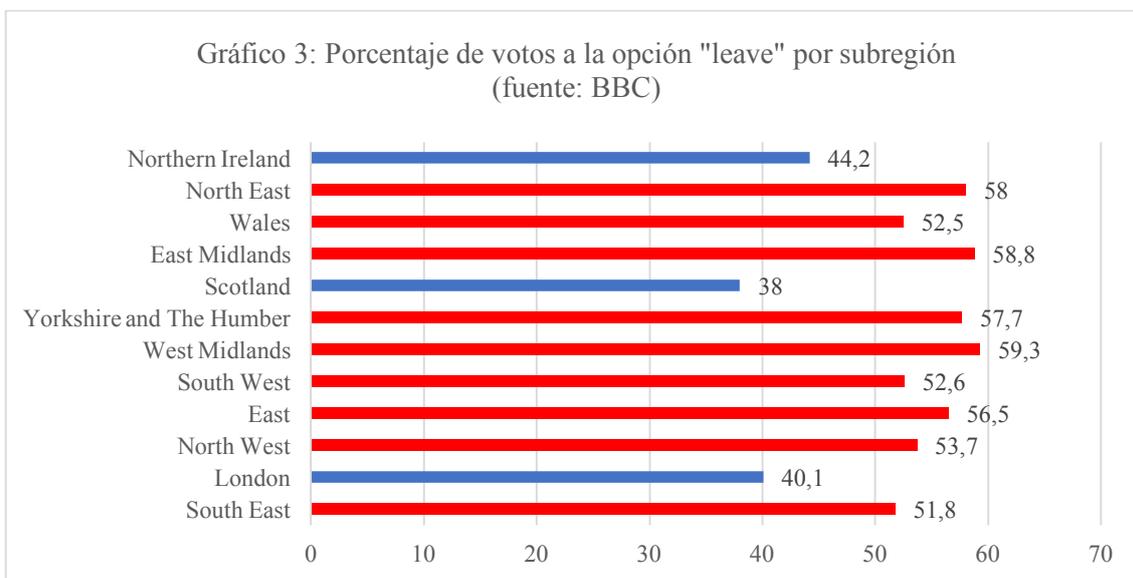
Procedamos ahora a observar el interés por subregión que suscitaron ambas opciones en la red. En primer lugar, se ha analizado el impacto de los temas de búsqueda “*Vote leave*” y “*vote remain*” en las diversas zonas del Reino Unido (gráfico 2). Si descodificamos los datos obtenidos desde Google Trends y los convertimos en un gráfico, quedaría de la siguiente manera¹⁸:

¹⁸ Google no ofrece datos geográficos más detallados que los que figuran en el gráfico.



Fuente: Elaboración propia con los datos de Google Trends

Si comparamos los datos superiores a los resultados electorales reales observaremos varios hechos relevantes. En el siguiente gráfico aparecen los resultados definitivos del referéndum divididos por subregión (gráfico 3). Se indican en rojo las zonas en las que la opción “leave” fue la más votada.



Fuente: Elaboración propia con los datos de BBC.

Los datos obtenidos del análisis de tendencias y los resultados finales coinciden en el caso de Escocia, donde se presenta un interés menor por el tema “vote leave” y con Gales, que sí presenta un relativo interés por la opción “leave”. En cambio, los datos muestran una clara disonancia en el caso de Irlanda del Norte. Google identifica a la región de Irlanda del Norte como muy interesada por el tema de búsqueda “vote leave”, algo que indica, si nos ceñimos a las explicaciones del buscador, un aumento desmesurado de las consultas en la red. Sin embargo, este aumento de las consultas no se vio reflejando en los resultados

finales. Irlanda del Norte fue una de las tres regiones con menos votos a favor de abandonar la Unión Europea. Este hecho es relevante, pues confirma que un mayor interés por un tema de búsqueda no va siempre acompañado de una tendencia de voto clara. En general, la tendencia de Google Trends exagera el tema de búsqueda “*vote leave*” e infravalora la opción “*vote remain*”. No está clara la naturaleza de esta tendencia. Como son muchos los factores que la pueden haber ocasionado, se han comparado los datos superiores con los obtenidos desde el contador de visitas de páginas de Wikipedia.

En cuanto refiere a las visitas de las páginas de Wikipedia, los resultados son un tanto más ambiguos. No existen, en Wikipedia, páginas específicas para las opciones “*remain*” y “*leave*” por tanto, es imposible utilizar la Wikipedia como un indicador fiable para determinar que opción fue la más solicitada. Sin embargo, sí podemos utilizar los datos de Wikipedia para analizar otro efecto importante del referéndum: la existencia de un mayor tráfico de red. Es decir, si durante la campaña del “Brexit” se dio un incremento del tráfico a las páginas de Wikipedia. Para hacerlo se utilizan como referencia tres entradas de la Wikipedia. Siendo estas “*Brexit*”¹⁹; “*United Kingdom European Union membership referendum, 2016*”²⁰; y “*Opinion polling for the United Kingdom European Union membership referendum*”²¹. Los lapsos de tiempo observados corresponden a las tres semanas anteriores al referéndum, desde el 1 de junio del 2016 hasta el 22 de junio del 2016 y coinciden con el caso anterior.

En primer lugar, si observamos la entrada de la Wikipedia que corresponde al “Brexit” la tendencia resulta clara (gráfico 4). Existe un aumento progresivo del número de visitas a la página. Cabe destacar que el grafico siguiente solamente corresponde al número de visitas de la página en idioma inglés.

¹⁹ Disponible en: <https://en.wikipedia.org/wiki/Brexit>

²⁰ Disponible en:

https://en.wikipedia.org/wiki/United_Kingdom_European_Union_membership_referendum,_2016

²¹ Disponible en:

https://en.wikipedia.org/wiki/Opinion_polling_for_the_United_Kingdom_European_Union_membership_referendum



Fuente: Elaboración propia con los datos de Wikipedia page views counter

En números absolutos, la página de Wikipedia fue vista un total de 44.164 veces en su versión inglesa durante las tres semanas previas a la consulta. No obstante, resulta curioso observar cómo, de nuevo, entre los días 16 y 18 de junio se produce un enorme declive en las visitas. No existen explicaciones fiables para justificar este descenso, pero la lógica nos lleva a pensar que esta caída está directamente relacionada con la muerte de la diputada Jo Cox.

Considero digno de mención destacar que la página española, titulada “Salida del Reino Unido de la Unión Europea” que redirige desde Brexit, fue vista más veces que la versión inglesa de la misma; un total de 49.397 veces²². Este hecho, si bien anecdótico, no ha sido analizado en profundidad pues considero que se desvía del objetivo principal del trabajo. No existen, a mi entender, suficientes potenciales participantes del referéndum en los territorios hispanoparlantes como para poder alterar los resultados del referéndum.

Si volvemos al caso que nos atañe, observemos a continuación el número de visitas a la página de Wikipedia titulada “*United Kingdom European Union membership referendum, 2016*” (Gráfico 5).

²² Fuente original:

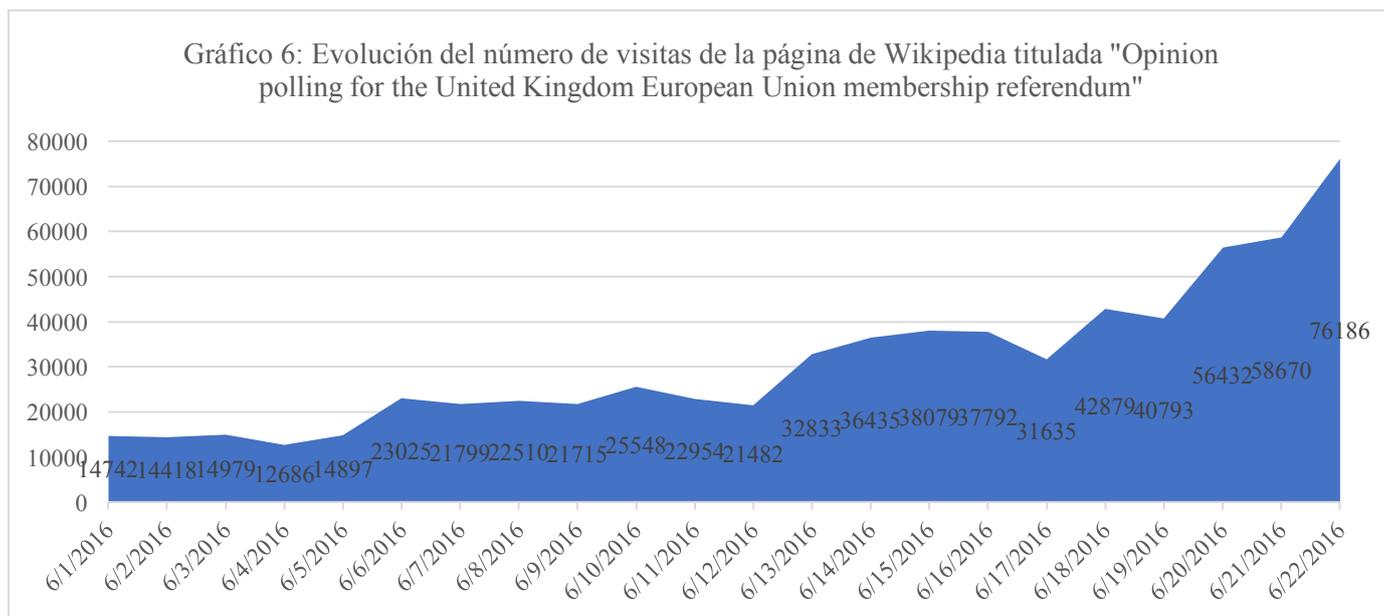
<https://tools.wmflabs.org/langviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2016-06-01&end=2016-06-22&sort=views&direction=1&view=list&page=Brexit>



Fuente: Elaboración propia con los datos de Wikipedia page views counter

En este segundo caso los datos son todavía más esclarecedores que en el anterior. En primer lugar, nos encontramos con un número de visitas muy superior a los del gráfico anterior. La página de Wikipedia titulada “*United Kingdom European Union membership referendum, 2016*” fue vista un total de 381.087 veces en su versión inglesa en las tres semanas anteriores al referéndum. Como se observa en la figura 3, el aumento es claro. Durante los primeros días del mes de junio, la página recibía unas 9000 visitas diarias, una cifra que aumentó considerablemente a medida que se acercaba el referéndum, hasta alcanzar las 51.879 visitas durante la jornada de reflexión del 22 de junio. De nuevo, cabe destacar que un abrupto descenso de las visitas se produce entre el 16 y el 18 de junio, coincidiendo con los datos de las dos graficas anteriores.

Veamos ahora el último caso, la página de Wikipedia titulada “*Opinion polling for the United Kingdom European Union membership referendum*” (gráfico 6).



Fuente: Elaboración propia con los datos de Wikipedia page views counter

Esta página de Wikipedia hace referencia únicamente a los métodos utilizados por la demoscopia del momento para determinar qué porcentaje de la población sería más favorable a votar por cada opción. Podríamos entenderlo, entonces, como un estudio demográfico aplicado al referéndum. Además, la entrada incluye los resultados de las diferentes encuestas que se habían realizado hasta entonces.

Cabe mencionar que la mayoría de las encuestas que aparecen en esta entrada ofrecían una ventaja ajustada a la opción “*remain*”, como hemos visto también en los casos anteriores. Sin embargo, el interés suscitado por esta entrada de la Wikipedia parece ser mayor del que podríamos esperar. En total, la versión inglesa de la página sumaba 682.313 visitas en su versión inglesa durante las tres semanas anteriores al referéndum. El aumento del número de visitas es, además, mucho más gradual que en los dos casos anteriores, pues podemos observar como durante las dos primeras semanas del mes de junio el número de visitas apenas varía, con un ligerísimo aumento entre el 12 y el 16 de junio. Las visitas aumentan a partir del 17 de junio, alcanzando el máximo de 76.186 visitas la jornada de reflexión del 22 de junio.

En general, podemos afirmar que el tráfico de red a páginas como Wikipedia aumentó considerablemente durante el proceso electoral del “Brexit”. Si bien estos datos no nos sirven para extrapolar una predicción válida, sí nos demuestran que el interés suscitado por el referéndum quedó reflejado en la red. Aumentó, por ende, el número de personas que querían conocer más datos sobre el referéndum. Este hecho lo podemos achacar a diversas razones: un mayor número de publicaciones sobre el tema, voluntad de conocer los últimos sondeos, interés general por las opciones, etc. Descubrir estas motivaciones, pero, no es el objeto de este trabajo y será necesaria una investigación más concreta.

El caso de Google Trends, sin embargo, entraña más dificultad. Si observamos las tendencias de búsqueda, Google acertó con su predicción del resultado final. Pero falló en algunos casos, en tanto que refiere el interés por subregión. Si bien es cierto que las tendencias de búsquedas en Google nos demuestran de forma clara que la opción del “*vote leave*” fue la más buscada, no podemos extraer conclusiones fehacientes de este hecho. Aun así, es digno mencionar que no resulta nada descabellado pensar que la mayoría de los usuarios que introducían los términos “*vote leave*” y “*vote remain*” en el buscador buscaban razones para votar a sendas opciones. De ser cierto este hecho, se podría demostrar que existe una correlación directa entre el número de personas que buscaban las razones y el número de votantes. No obstante, hará falta analizar los dos casos siguientes para extraer conclusiones definitivas.

5.2 El caso estadounidense: Elecciones presidenciales de noviembre del 2016.

Las elecciones de noviembre del 2016 en Estados Unidos de América tuvieron una relevancia enorme en el panorama político mundial. Tras dos administraciones del presidente Barack Obama, Estados Unidos se preparaba para afrontar una nueva etapa en su historia.

Los ocho años de presidencia de Obama habían empezado con gran entusiasmo; sin embargo, los últimos dos años de su presidencia dejaron muchas promesas incumplidas: continuas tensiones raciales no resueltas; esporádicas masacres con unas armas de fuego que todavía no habían sido reguladas; fallos en las políticas de exterior en países como Siria o Libia; el incumplido cierre de la prisión de Guantánamo; etc. El descontento generalizado con su administración contrastaba con el optimismo inicial de su reelección en el 2012. El candidato que heredara el poder tendría que enfrentarse a grandes retos, no solo en Estados Unidos sino también en el mundo entero.

El sistema político estadounidense es presidencialista, algo que se refleja en la manera en la que funcionan los procesos electorales en este país. El sistema presidencialista, por regla general, provoca que las campañas electorales giren en torno a los candidatos en cuestión casi de forma exclusiva. Esto queda patente en el caso americano, donde todos los eslóganes de campaña hacen referencia al nombre o/y apellido del candidato. De esta manera, tras las sendas nominaciones a los candidatos del partido demócrata y el republicano, los eslóganes de campaña fueron los siguientes:

Partido Demócrata (DP)	Hillary Clinton	Hillary for America	Clinton / Kaine Hillary 2016
Partido Republicano (GOP)	Donald J. Trump	Make America Great Again!	Trump / Pence Trump 2016

Los datos superiores se muestran únicamente para enseñar que el carácter individualista de las elecciones presidenciales en Estados Unidos facilita el filtrado de información. De esta forma, cuando analizamos las tendencias de búsqueda de un candidato para un determinado periodo de tiempo, podemos utilizar su nombre y/o apellidos para filtrar la búsqueda. En el caso de sistemas políticos en los que se potencia la figura del partido, como el caso británico, nos encontramos con un abanico de posibilidades mucho más amplio. En el caso inglés, por ejemplo, podemos referirnos al partido obrero de varias maneras: “*Labour*”; “*Labour party*”; “*LP*” o simplemente por el nombre de su candidato: “*Corbyn*” o “*Jeremy Corbyn*”. Esta gran variedad de posibles términos de búsqueda dificulta el filtrado de información. Esto es así porque el buscador hegemónico de la red, Google, no redirige todos los términos de búsqueda anteriores a los mismos resultados. De esta manera, obtendremos resultados distintos si buscamos “*Labour*” que si

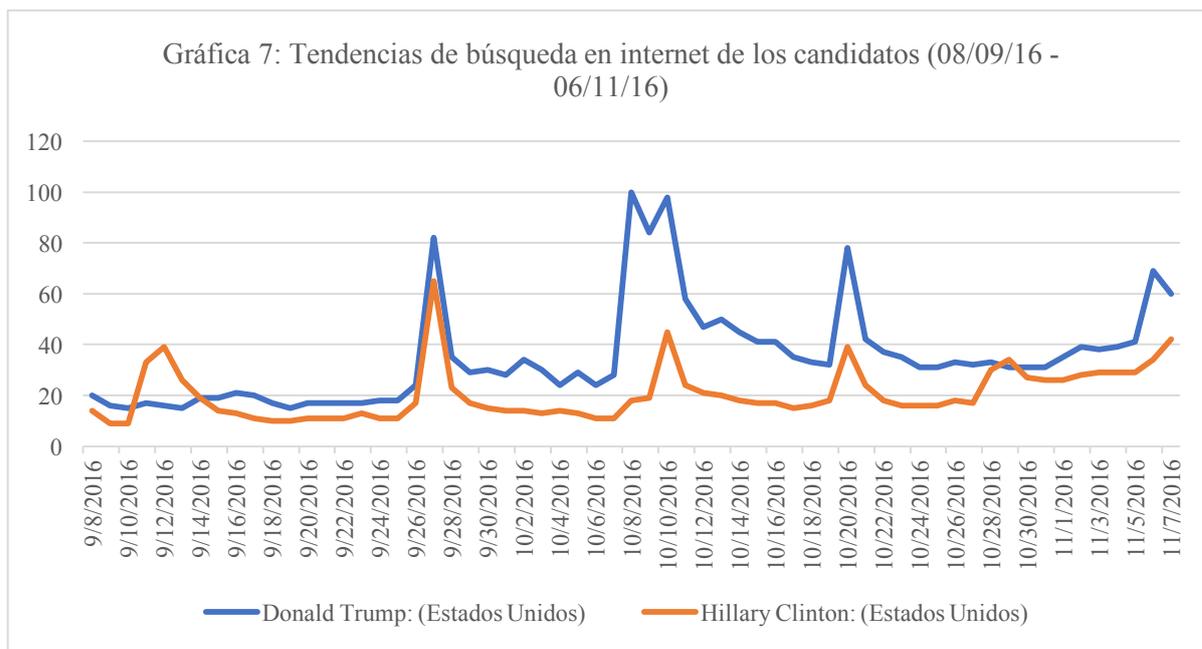
escribimos “*Labour Party*” en el buscador. El sistema de búsqueda de Google identifica de forma distinta los términos “*Labour*” y “*Labour Party*”, identificando el primero como un tema de búsqueda genérico y el segundo como un término de búsqueda concreto, referido directamente al partido obrero británico²³. Todo esto provoca que las tendencias de búsqueda varíen en función de los términos introducidos en el buscador.

No ocurre lo mismo en el caso americano, donde los términos de búsqueda coinciden con nombres concretos de personas. De esta forma, cuando introducimos el nombre “Donald Trump” Google nos devuelve los resultados de la persona asociada a este nombre, en nuestro caso, el 45º presidente de los Estados Unidos de América. Existen, en definitiva, menos formas de referirse a una persona en concreto que a un partido político, y eso reduce considerablemente el volumen de información por filtrar.

Si volvemos al caso que nos atañe, las elecciones estadounidenses del 2016 tuvieron un grandísimo impacto en la red, con ambos candidatos lanzando mensajes continuos en las redes sociales. Este componente “social” que empezó a desarrollarse para la campaña presidencial de Obama en el 2008 (Bloggin Zenith, 2013), ha ido tomando mucha fuerza en los últimos años y hoy es la piedra angular de cualquier campaña política que se precie. No fue distinto en el caso de las elecciones de noviembre del 2016, donde ambos candidatos utilizaron muy frecuentemente las redes sociales; especialmente el candidato republicano Donald Trump, que incluso fue apodado el presidente del Tuit (Juez, 2016).

Esta fuerte presencia en las redes fue un factor decisivo para la victoria del candidato republicano, pues todas sus afirmaciones en la red, algunas de ellas polémicas, incrementaron enormemente el volumen de búsquedas en internet. Lo mismo ocurrió con los debates televisados, que responden a un gran incremento del interés general por ambos candidatos. Sin embargo, hay un hecho muy relevante que demuestra la desconexión entre la percepción de los medios de comunicación y la realidad en la red. Veamos, a continuación, las tendencias de búsqueda en Google durante las tres semanas anterior a las elecciones (gráfico 7):

²³ Más información sobre las diferencias entre tema y termino de búsqueda se encuentra en el apartado “Metodología” de este trabajo y en la sección de preguntas frecuentes de Google Trends: <https://support.google.com/trends/answer/4359550>



Fuente: Elaboración propia con los datos de Google Trends

Observamos en el gráfico como el interés en la red parece ser siempre superior por Donald Trump durante los dos debates televisados que se realizaron, tanto en el del 27 de setiembre como en el del 10 de octubre. Sin embargo, sin nos fijamos en las noticias publicadas por medios generalistas a raíz del debate, todos coinciden en nombrar a Hillary Clinton la gran triunfadora. Aquí algunos ejemplos.

- *Consensus: Clinton wins debate – but Trump is far from finished*; The Guardian; 27/09/16²⁴
- *Post-debate poll: Hillary Clinton takes round one*; CNN; 27/09/16²⁵
- *Presidential debate: Hillary Clinton wins second head-to-head, poll suggests*; Independent; 10/10/16²⁶
- *Clinton wins debate, but Trump exceeds expectations*; CNN; 10/10/16²⁷
- *Poll: Hillary Clinton won the second debate*; Politico; 11/10/16²⁸

Vemos pues, que, si bien es cierto que la victoria de Clinton en los dos debates presidenciales obtuvo cierto consenso por parte del público, esta percepción no corresponde al interés real suscitado por ambos candidatos en internet.

Traslademos ahora el mismo ejercicio a la página de Wikipedia. Durante el mismo lapso

²⁴ Disponible en: <https://www.theguardian.com/us-news/2016/sep/27/debate-polls-reaction-hillary-clinton-donald-trump>

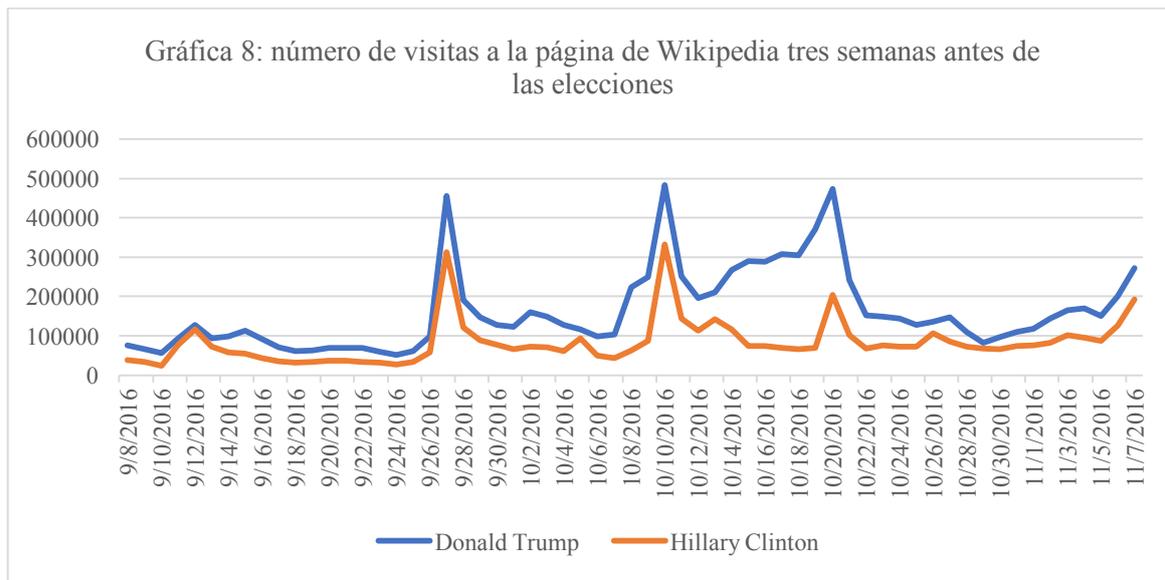
²⁵ Disponible en: <http://edition.cnn.com/2016/09/27/politics/hillary-clinton-donald-trump-debate-poll/>

²⁶ Disponible en: <http://www.independent.co.uk/news/world/americas/presidential-debate-poll-winner-hillary-clinton-donald-trump-a7353391.html>

²⁷ Disponible en: <http://edition.cnn.com/2016/10/09/politics/clinton-wins-debate-but-trump-exceeds-expectations/>

²⁸ Disponible en: <http://www.politico.com/story/2016/10/clinton-trump-debate-poll-229581>

de tiempo, desde el 8 de setiembre hasta el siete de noviembre, los resultados son parecidos. Durante este periodo de tiempo, la página de Wikipedia de Donald J. Trump obtuvo un total de 9.920.187²⁹ millones de visitas, únicamente en su versión inglesa, con un promedio diario de 162.626 visitas al día. Por el contrario, la página de Wikipedia de la candidata Hillary Clinton fue visitada 5.101.041³⁰ millones de veces, con una media de 83.624 visitas al día. Estos datos, se reflejan también en el eje temporal, con Donald Trump siempre a la cabeza, tal y como muestra la gráfica 8.



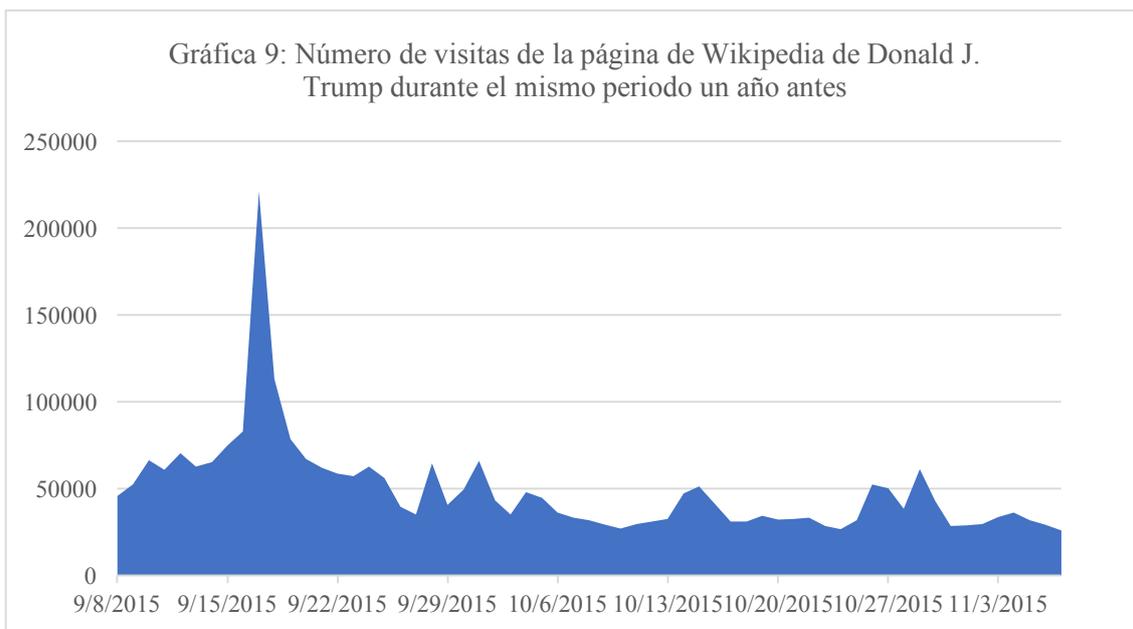
Fuente: Elaboración propia con los datos de Wikipedia page views counter

De nuevo, Donald Trump obtuvo más visitas en su página de Wikipedia tras los debates que Hillary Clinton. Una de las razones por las que el número de visitas a la página de Wikipedia de un candidato resulta tan esclarecedor, es porque implican un interés por el candidato en cuestión. En este punto, las tendencias en Wikipedia resultan mucho más fiables que las tendencias de búsqueda genéricas en buscadores como Google. El motivo es claro: una persona que accede a la página de Wikipedia de una persona o tema, por definición, busca información sobre aquello que ha buscado. No ocurre lo mismo con Google, donde una búsqueda puede corresponder tanto al interés que suscita el tema, la búsqueda de una noticia en concreto o simplemente, una consulta gramatical (Gonzalo, 2013). Este hecho queda probado si comparamos el número de visitas de ambos candidatos en la figura 8 con el número de visitas a las mismas páginas durante el mismo lapso de tiempo un año antes (gráficas 9 y 10).

²⁹ Información extraída de un contabilizador de visitas de Wikipedia convertido en interfaz gráfica.

Disponible en: https://tools.wmflabs.org/langviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2016-09-08&end=2016-11-07&sort=views&direction=1&view=list&page=Donald_Trump

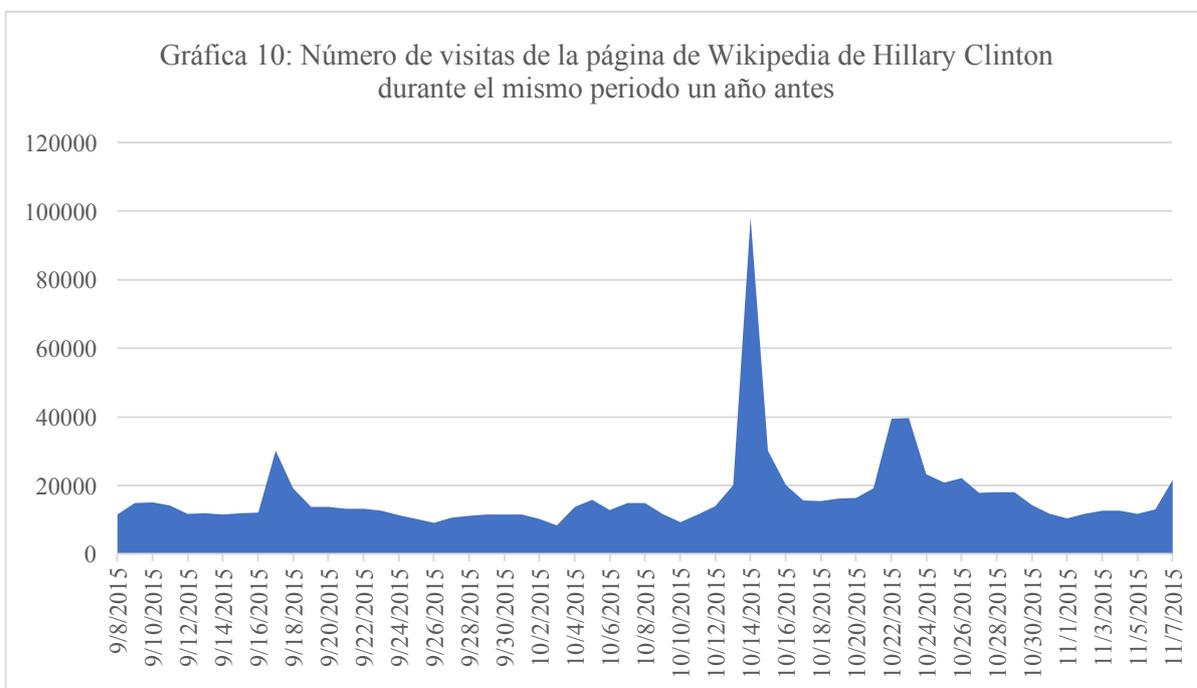
³⁰ Ídem.



Fuente: Elaboración propia con los datos de Wikipedia page views counter

La página inglesa de Donald Trump fue visitada 2.990.237 de veces entre el 8 de setiembre del 2015 y el 7 de noviembre del mismo año.

La página de Hillary Clinton también muestra un índice de visitas muy inferior que durante el periodo electoral. En total, entre el 8 de setiembre del 2015 y el 7 de noviembre del mismo año, la página de la, por entonces, Secretaria de Estado, Hillary Clinton, fue vista un total de 1.025.219 veces, tal y como indica la figura 10:

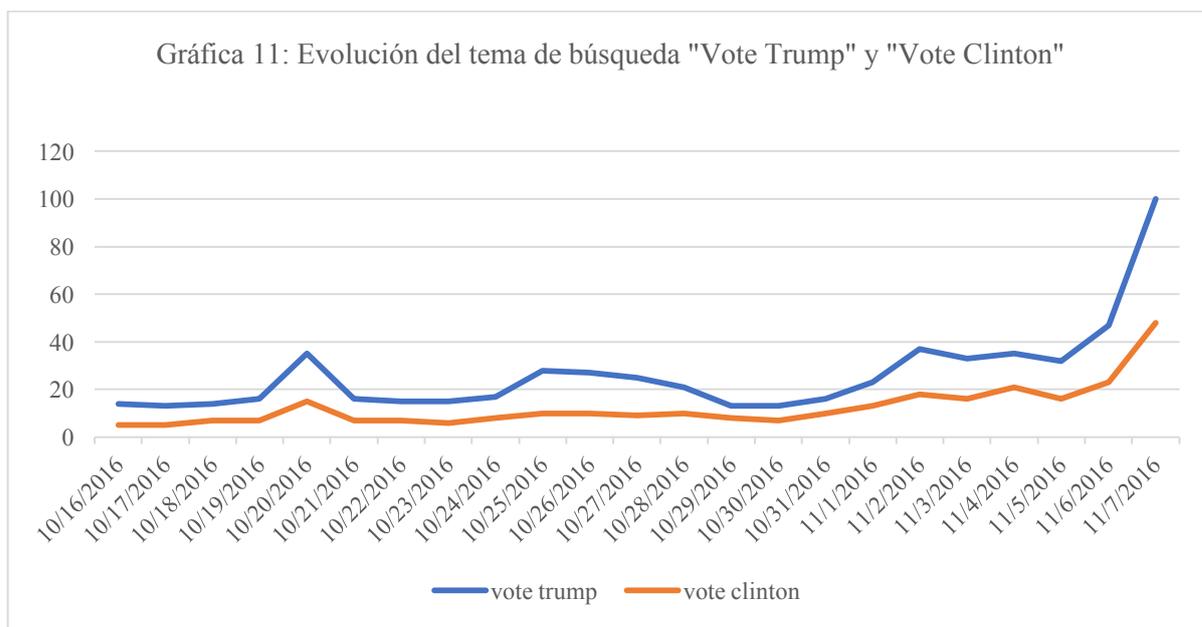


Fuente: Elaboración propia con los datos de Wikipedia page views counter

Observamos, por tanto, que el tráfico en la red hacia páginas como Wikipedia crece

enormemente antes de periodos electorales. Los datos muestran que el número de visitas a las páginas de Wikipedia de ambos candidatos crecieron un 331 % en el caso de Donald J. Trump y un 497% en el caso de Hillary Clinton. Este aumento de las búsquedas en Wikipedia parece indicar un mayor interés por el sujeto; efecto esperado durante un proceso electoral.

En las gráficas superiores, por tanto, hemos visto como las búsquedas en internet fueron especialmente dirigidas a Donald J. Trump y su campaña, algo que posteriormente coincidió con los resultados electorales. Naturalmente, no podemos simplemente suponer que los datos obtenidos a partir del análisis de tendencias nos sirvan como un indicador electoral válido. Por ese motivo vamos a filtrar las tendencias de los términos de búsqueda “Donald J. Trump” y “Hillary Clinton” comparándolos con los temas de búsqueda “Vote Trump” y “Vote Clinton”³¹. Estos últimos nos darán una noción más específica de la cantidad de personas que buscaron motivos por los que votar a sendos candidatos. Además, estos datos nos dan un claro indicador de quien despierta más interés en la población participe en la red (gráfica 11).

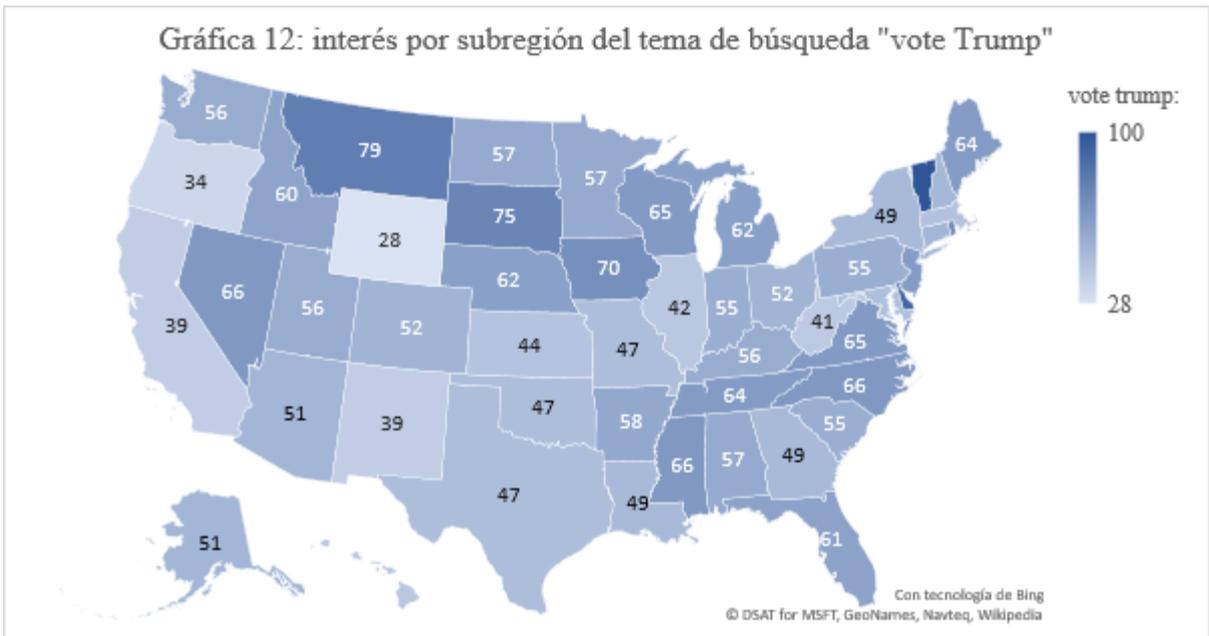


Fuente: Elaboración propia con los datos de Google Trends

Observamos en el gráfico superior (gráfica 11), que el tema de búsqueda “Vote Trump” lideró las tendencias en Google durante las tres semanas previas a las elecciones. Este hecho es relevante pues coincide con las tendencias observadas en el gráfico anterior.

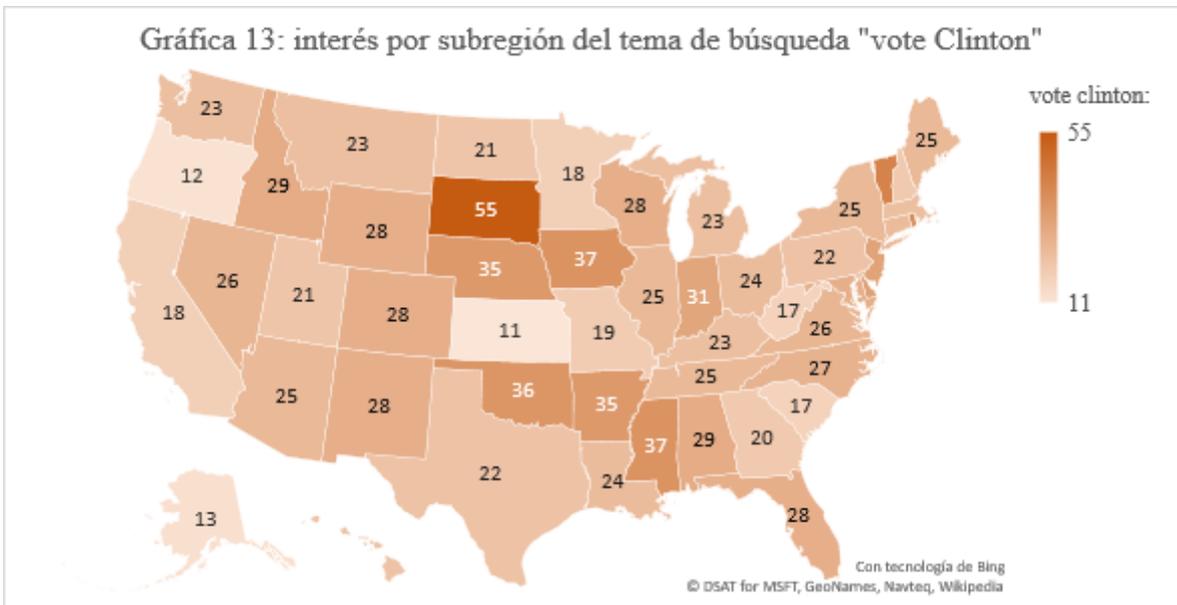
Procedamos ahora a analizar el interés por subregión de los temas de búsqueda “Vote Trump” y “Vote Clinton”. Si decodificamos los datos obtenidos a partir de Google Trends el mapa estadounidense quedaría de la siguiente manera (Gráfica 12).

³¹ Más información sobre las diferencias entre tema de búsqueda y término de búsqueda se encuentra en la sección Metodología de este trabajo.



Fuente: Elaboración propia con los datos de Google Trends

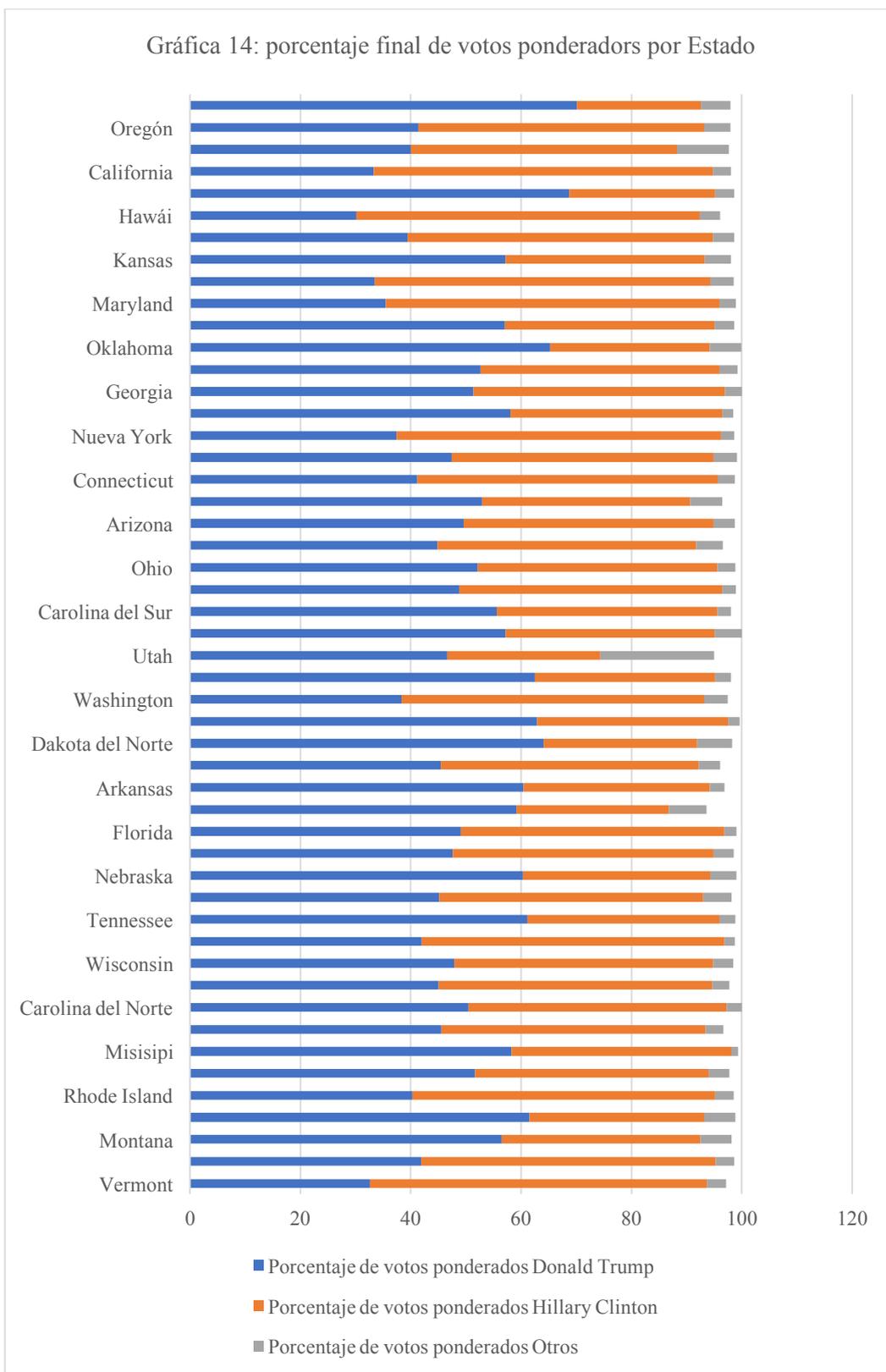
En el mapa superior se aprecia el número de búsquedas ponderadas por Estado. En contraste, la actividad bajo el tema de búsqueda “*vote Clinton*” fue mucho más moderada (gráfica 13).



Fuente: Elaboración propia con los datos de Google Trends

En general, los datos reflejan que hubo una mayor actividad bajo el tema de búsqueda “*vote Trump*” y esta quedó más repartida. En ningún Estado, a excepción de Wyoming, el tema de búsqueda “*vote Clinton*” superó al tema “*vote Trump*”. Ahora comparemos estos datos con los resultados oficiales de las elecciones; extraídas de página web del

periódico “El País”³² (gráfica 14).



Fuente: Elaboración propia con los datos de El País

³² Disponible en: <http://elpais.com/especiales/2016/elecciones-eeuu/resultados-electorales/>

En general, los datos coinciden relativamente bien con los obtenidos a partir de las tendencias de Google Trends; donde hubo una mayor actividad del tema de búsqueda “*vote Trump*” también hubo un mayor porcentaje de electores que votaron en su favor. Sin embargo, hay una serie de flagrantes discrepancias que merecen ser tratadas. En primer lugar, Wyoming, el Estado donde el tema de búsqueda “*vote Trump*” fue más débil, fue, paradójicamente, donde el candidato republicano obtuvo los mejores resultados de todo el país. Algo que resulta paradójico si nos basamos en los datos de Google Trends. Y, con el mismo resultado, pero en el lado opuesto del espectro, se encuentra el Estado de Vermont, donde el tema de búsqueda “*vote Trump*” llegó a los 100 puntos, indicando un aumento desmesurado de las búsquedas. No obstante, en los resultados finales obtuvo solamente un 32,6 por ciento de votos para el candidato republicano, el resultado más bajo de todo el país. No existe una explicación clara a esta paradoja. El estado de Vermont es considerado uno de los más progresistas del país y ningún candidato republicano ha superado el 40% de los votos desde que lo hiciera Georges W. Bush en el año 2000. Su alto índice de búsquedas por candidatos republicanos contrasta con los resultados que se suelen obtener durante las elecciones.

El caso de Hillary Clinton presenta discrepancias similares. Si bien es cierto que en general la actividad del tema de búsqueda “*vote Clinton*” fue moderada, algunos Estados presentan datos relevantes. De nuevo, nos encontramos con el Estado de Wyoming, donde las búsquedas del tema “*vote Clinton*” llegaron a los 100 puntos según Google Trends. Sin embargo, estos datos, de nuevo, no coinciden con el resultado final. En dicho Estado, Clinton solo obtuvo un 22,5% de los votos, el mínimo de todo el país.

No existen razones claras que expliquen la discrepancia que presentan algunos Estados como Wyoming o Vermont entre los datos de Google Trends y los resultados finales. Sin embargo, sí podemos afirmar que hubo una mayor actividad en la red bajo el tema de búsqueda “*vote Trump*”. Este hecho es relevante porque fue infravalorado por la demoscopia tradicional. Los datos en la red, tal y como afirmó el ingeniero estadounidense de origen indio, Sanjiv Rai, en un informe para la CNBC, mostraban un claro dominio del candidato republicano en la red. De hecho, según el mismo, Donald Trump era más popular en la red que Barack Obama en el 2008 (Kharpal, 2016). Sanjiv Rai incluso llegó a afirmar: *"If Trump loses, it will defy the data trend for the first time in the last 12 years since Internet engagement began in full earnest."* (Kharpal, 2016)

Incluso el sistema de Inteligencia Artificial, MogIA, desarrollado en el 2004 por Sanjiv Rai predijo de forma correcta las últimas 4 elecciones estadounidenses (Contreras, 2016). Los datos obtenidos por el análisis de Big Data demostraron, en casi todas las proyecciones realizadas, que la victoria de Donald Trump era una opción fehaciente. Algo que parece coincidir con los análisis aquí expuestos.

5.3 El caso francés: elecciones presidenciales de abril-mayo 2017

La forma de gobierno francesa es semipresidencialista, y su sistema electoral corresponde con un modelo arquetípico de sistema mayoritario a dos vueltas. En Francia, obtiene la presidencia el candidato más votado, pero con el añadido de que para obtener la presidencia se necesitan superar, por regla general, dos vueltas, también llamadas balotaje³³.

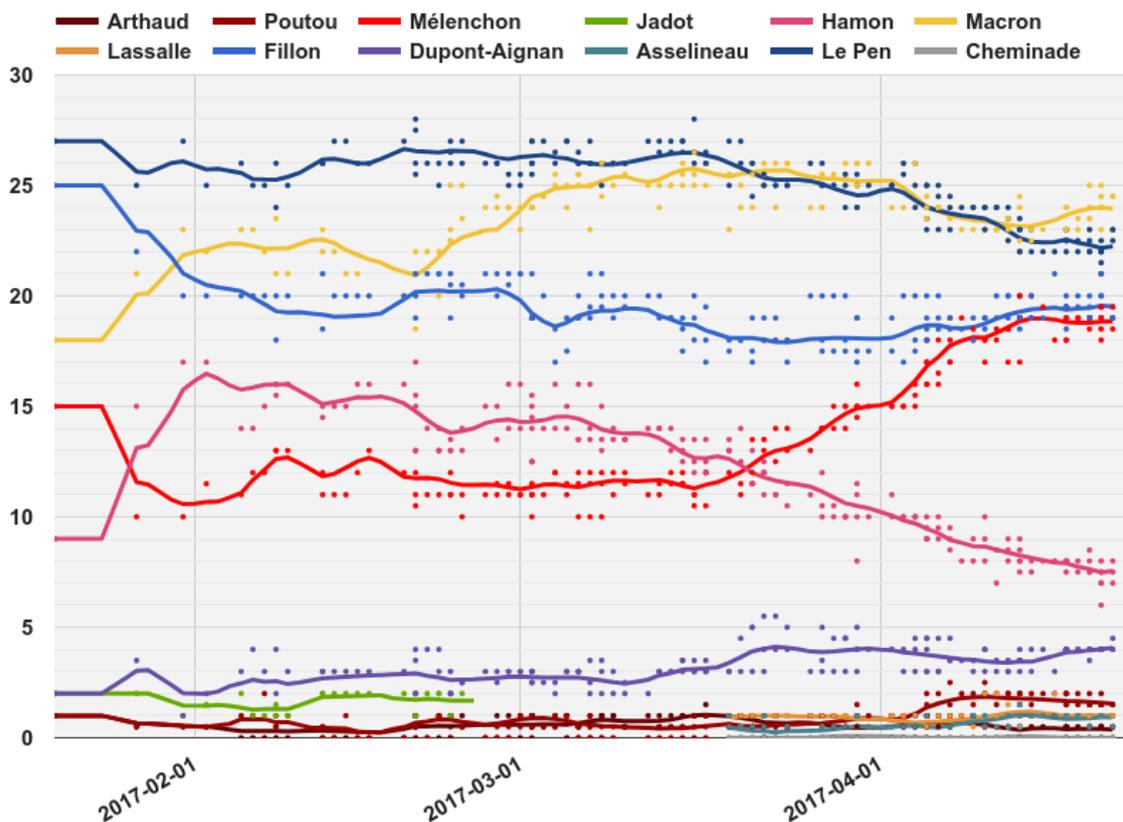
Estos datos son relevantes para el presente trabajo en tanto que demuestran que, para predecir las elecciones francesas a través de métodos de análisis de Big Data, es necesario realizar dos operaciones independientes. Por un lado, se debe analizar la popularidad en la red de todos los candidatos presidenciables durante las semanas previas a la primera vuelta. Estos datos no son significativos en cuanto refiere el resultado final, pues es posible que el candidato más popular en la primera vuelta no sea el elegido durante la segunda. Los datos de la primera vuelta, por tanto, son orientativos y sirven para poner a prueba la metodología. En cambio, la segunda vuelta nos presenta resultados filtrados y por tanto más precisos, pues la decisión es únicamente entre dos candidatos. No obstante, para predecir la segunda vuelta disponemos únicamente de los datos correspondientes a dos semanas, pues es ese el lapso de tiempo entre la primera vuelta y la segunda.

Las elecciones presidenciales francesas del 2017 fueron decisivas para todo el continente europeo. En la prensa, la batalla por el Eliseo fue presentada como una lucha entre el europeísmo y el sentido común contra el extremismo radical. El escenario político fue, como poco, peculiar. Por primera vez en la Quinta República el presidente saliente, el socialista François Hollande, decidió no presentarse a su reelección. Algo que venía anunciándose desde meses anteriores y que quedó reflejado en la poca popularidad de la que gozó Hollande en sus últimos años de presidente. De hecho, François Hollande ostenta el infame título de presidente más impopular de la república (Valderrama, 2017). Esta noción entre la población francesa responde a varios hechos: la crisis migratoria europea, las políticas neoliberales implementadas, muy alejadas de su programa político inicial y, por último, el duro azote terrorista que convirtió a Francia en el país con más muertos por terrorismo yihadista del continente. La primera vuelta de las elecciones francesas aglutinó a varios candidatos de trasfondos muy dispares. Entre los favoritos, François Fillón, candidato para el partido republicano; Benoît Hamon, el solitario candidato del impopular partido socialista francés; Jean-Luc Mélenchon, el representante del movimiento popular “Francia Insumisa” y muy activo en la red; Emmanuel Macron, candidato independiente de la derecha moderada con su partido “En Marche!”, y, por último, Marine Le Pen, hija predilecta del clan Le Pen y líder del partido ultraderechista “Front National”.

³³ Es cierto que existe la posibilidad de ser elegido en la primera vuelta, pero sería necesaria una mayoría absoluta, algo que no suele suceder.

5.3.1 La primera vuelta:

La primera vuelta de las elecciones francesas fue intrigante como poco. Las encuestas no parecían tener claro que dos candidatos pasarían a la segunda vuelta, pero sí sabían que no habría mayoría absoluta³⁴. Si observamos la media entre las principales encuestas mayoritarias de Francia nos encontramos con el siguiente gráfico.

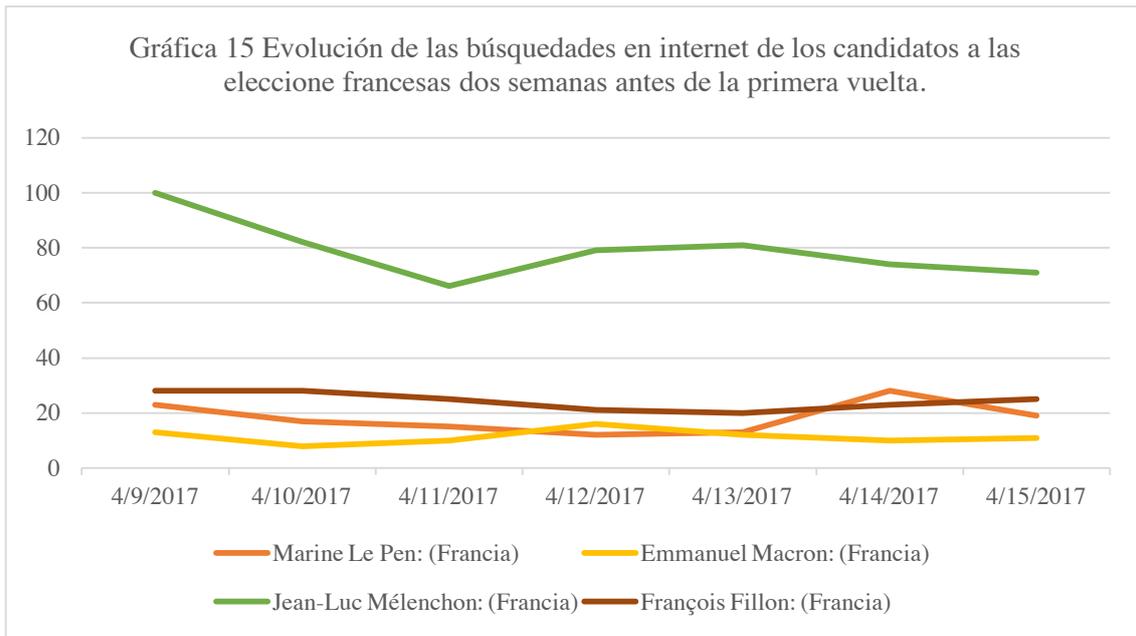


Fuente: Wikipedia, (2017)

En el gráfico superior observamos como los cuatro candidatos favoritos muestran predicciones muy parecidas. Todas entre el 20% y el 25% de los votos. Destacan Emmanuel Macron y Marine Le Pen con poca diferencia entre ambos. Por primera vez, las elecciones presentaban un escenario inédito en para Francia. Observemos ahora los mismos candidatos en el análisis de tendencia de búsqueda en internet dos semanas antes de la primera vuelta³⁵ (gráfica 15).

³⁴ Más datos sobre las predicciones de la primera vuelta en: *Opinion polling for the French presidential election, 2017*. Wikipedia. 2017. Disponible en: https://en.wikipedia.org/wiki/Opinion_polling_for_the_French_presidential_election,_2017

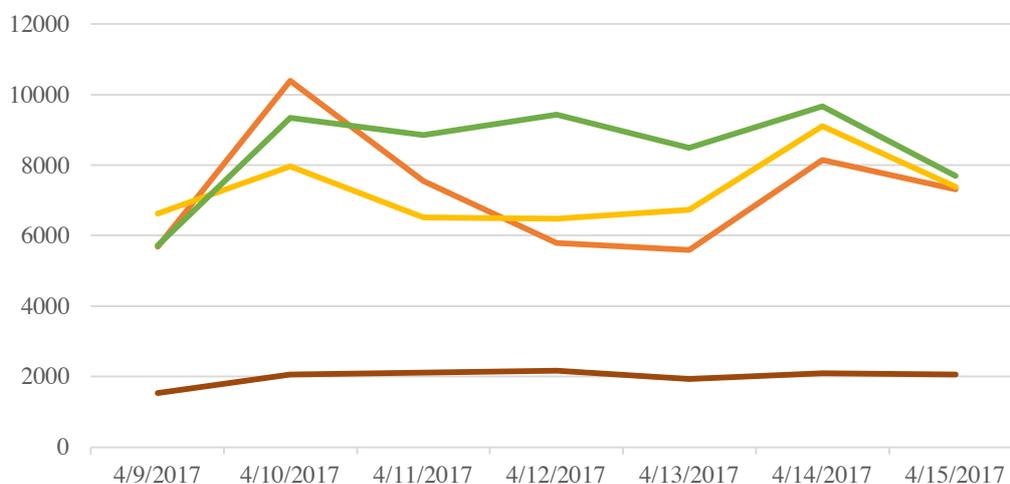
³⁵ En aras de la simplicidad solo se muestran los cuatro candidatos con más posibilidades de ser seleccionados: Marine Le Pen, François Filló, Jean-Luc Melénchon y Emmanuel Macron.



Fuente: Elaboración propia con los datos de Google Trends

Los datos superiores corresponden a la semana del 9 al 15 de abril del 2017. El candidato más buscado en Google fue Jean-Luc Melénchon, con una gran diferencia respecto a los demás. En comparación con las encuestas tradicionales, Google parece sobre-representar al candidato de la Francia Insumisa. Sin embargo, sitúa a los otros tres candidatos en posiciones muy parecida. Todas cercanas a los 20 puntos en las búsquedas. A continuación, se muestran los números de visitas de la página de Wikipedia de los candidatos (gráfica 17):

Gráfica17: Evolución del número de visitas a las páginas de Wikipedia de los candidatos dos semanas antes de la primera vuelta.



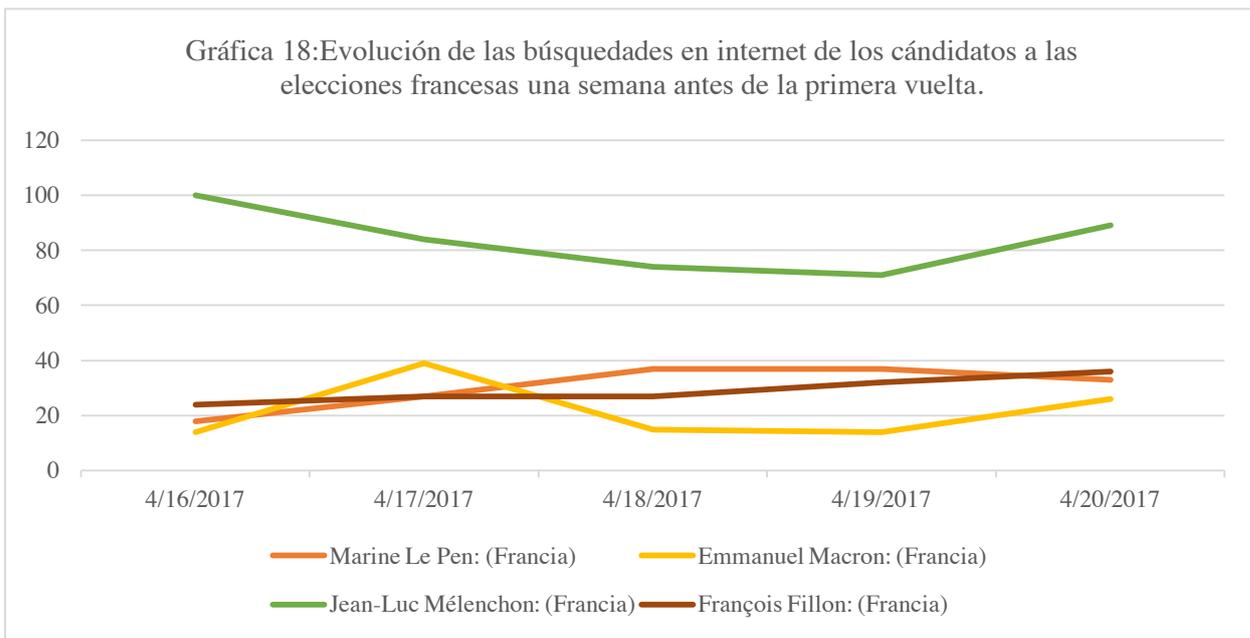
	4/9/2017	4/10/2017	4/11/2017	4/12/2017	4/13/2017	4/14/2017	4/15/2017
Marine Le Pen	5687	10402	7558	5792	5594	8141	7322
Emmanuel Macron	6634	7962	6519	6478	6728	9106	7391
Jean-Luc Mélenchon	5724	9341	8859	9432	8486	9668	7700
François Fillon	1524	2049	2118	2167	1937	2089	2056

Marine Le Pen Emmanuel Macron Jean-Luc Mélenchon François Fillon

Fuente: Elaboración propia con los datos de Wikipedia page views counter

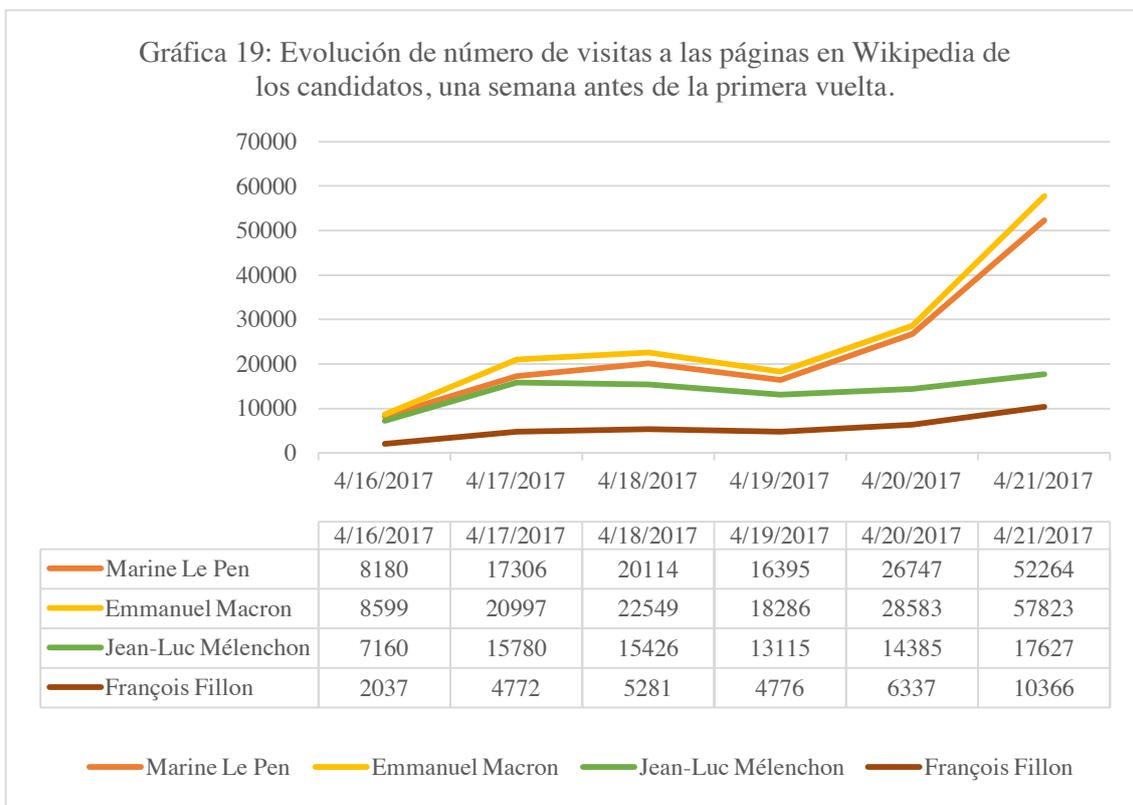
Los datos de la Wikipedia parecen ser más realistas cuando los comparamos a los obtenidos a partir de las encuestas. Los cuatro candidatos recibieron casi el mismo número de visitas a su página de Wikipedia, y no se muestra ninguna sobre representación que llame la atención. Si atendemos al número de visitas en Wikipedia, podemos afirmar que no hay ninguna certeza clara para la primera vuelta.

Sin embargo, es en la semana anterior a la primera vuelta cuando los datos empiezan a discrepar. En primer lugar, tenemos el número de búsquedas en Google, que muestran, de nuevo, una clara mayoría para el candidato de izquierdas Jean-Luc Mélenchon (gráfica 18):



Fuente: Elaboración propia con los datos de Google Trends

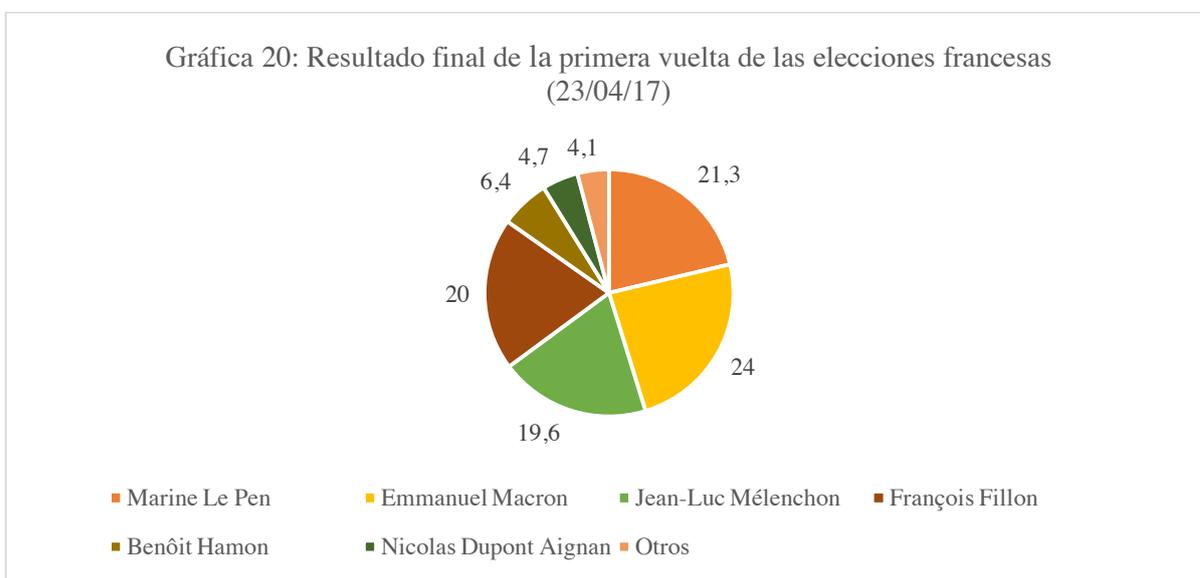
Además, Le Pen y Fillón parecen empatados y Macron ha experimentado un ligero descenso en las búsquedas. Los datos, por tanto, no parecen coincidir con ninguna de las encuestas realizadas. Por otro lado, tenemos a la Wikipedia, cuyo número de visitas nos muestran proyecciones diferentes (gráfica 19):



Fuente: Elaboración propia con los datos de Wikipedia page views counter

El número de visitas a la página de Wikipedia nos muestra unos resultados diferentes. El indicador de tendencias en Wikipedia nos muestra como las páginas de los candidatos Fillon y Melénchon experimentaron una bajada importante durante la semana anterior a la primera vuelta. En cambio, los candidatos Marine Le Pen y Emmanuel Macron, preferidos de las encuestas para pasar a la segunda vuelta, se sitúan por encima. La diferencia empieza a agravarse a partir de día 20 de abril, cuando el número de visitas diarias se incrementó hasta llegar a 30.000 visitas más al día (21/04/17). En cuanto respecta a los dos candidatos favoritos, Emmanuel Macron presenta una leve ventaja respecto a Le Pen, con 5559 visitas más a su página de Wikipedia el día 21 de abril, dos días antes de la primera vuelta.

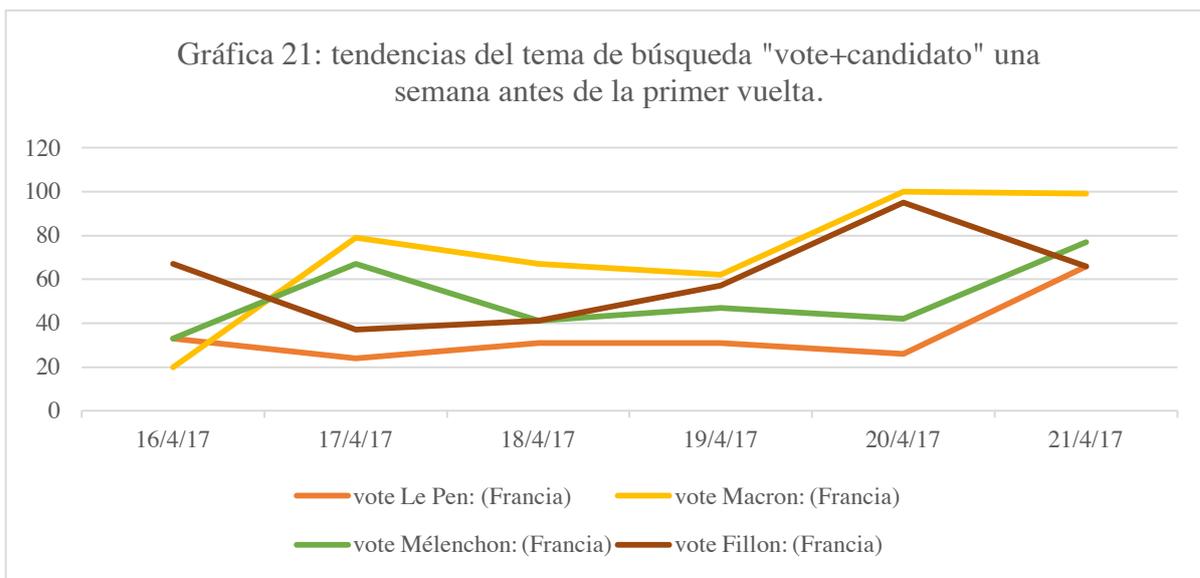
Los datos de Wikipedia llaman al optimismo. El número de visitas a las páginas de los candidatos corresponde, casi con exactitud, a los resultados obtenidos por los candidatos en la primera vuelta de las elecciones. Veamos a continuación el resultado (gráfica 20).



Fuente: Elaboración propia con los datos de 20 Minutos

Apreciamos en el gráfico como el candidato con más votos, Emmanuel Macron, fue también el más visitado en el contador de visitas de la página de Wikipedia. El mismo contador de visitas también acertó la segunda posición, la de Marine Le Pen, que obtuvo un 21,3 % de los votos, tan solo un poco por debajo de los votos de Macron. Una tendencia que también observamos anteriormente en la gráfica 19.

En general, podemos afirmar que los datos obtenidos a través del contador de visitas de Wikipedia fueron más fiables a la hora de predecir el desenlace electoral que los datos obtenidos por Google Trends. Sin embargo, para seguir con la misma metodología utilizada en los dos casos anteriores, se han analizado también los datos de Google Trends utilizando el tema de búsqueda “vote + nombre del candidato”. Los resultados de este análisis se muestran en la gráfica 21.

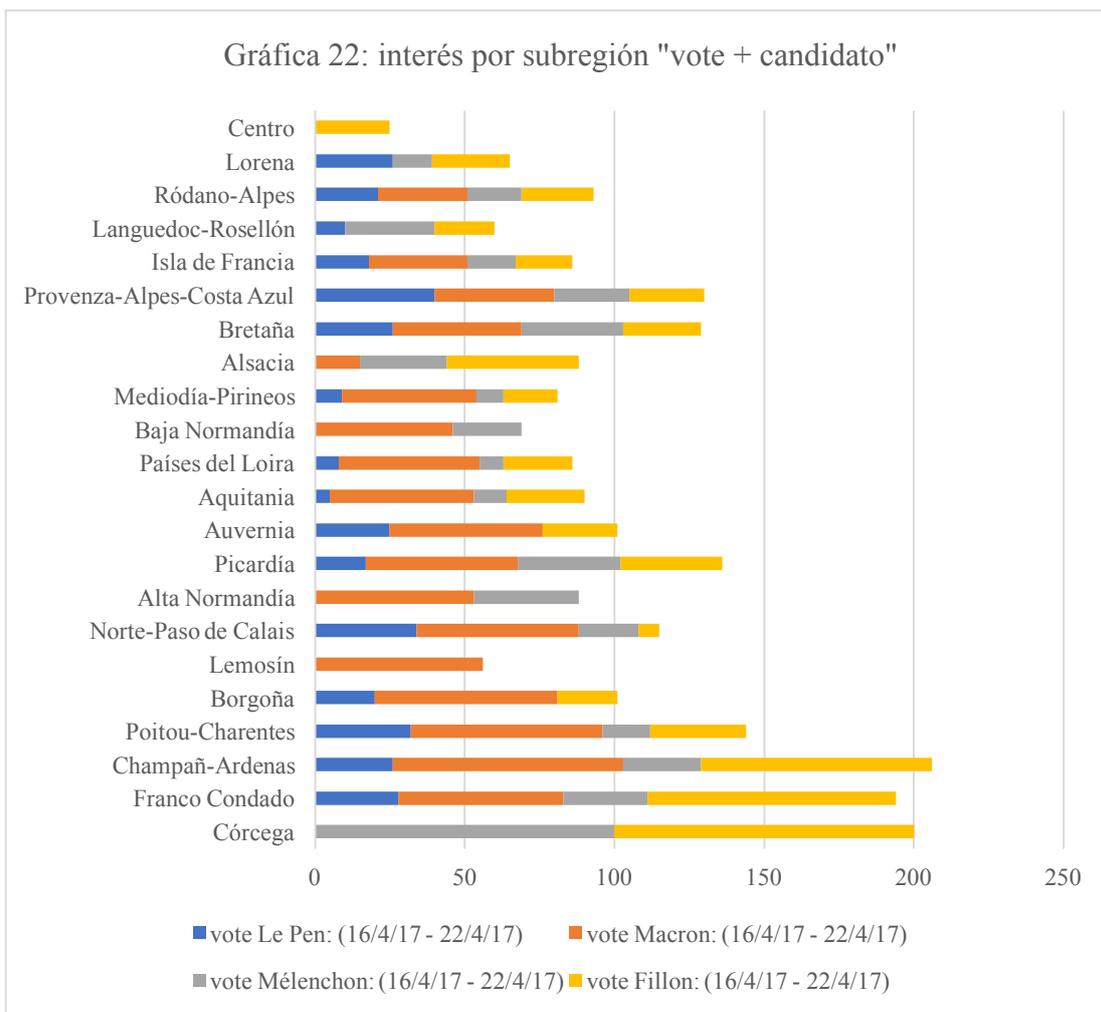


Fuente: Elaboración propia con los datos de Google Trends

Si aplicamos la fórmula “vote + nombre del candidato”, los datos que obtenemos son muy distintos. Si en la gráfica 18 observamos como el candidato más buscado en internet fue Jean-Luc Mélenchon, aplicando el filtro anterior, la delantera la lleva el tema de búsqueda “*vote Macron*”. No obstante, podemos observar como de cerca le sigue el tema “*vote Fillon*”, una lectura que no encaja en los resultados finales, aunque, como apreciamos en las tendencias de los días posteriores, las búsquedas cayeron drásticamente. En segundo lugar, el día 21 de abril, un día antes de la jornada de reflexión, se encuentra el tema “*vote Mélenchon*”, seguido de cerca de “*vote Fillon*” y “*vote Le Pen*”, que alcanzaron un mismo volumen de búsquedas.

Podemos afirmar que, aplicando el filtro adecuado, “vote + nombre de candidato”, Google es capaz de ofrecernos una lectura más correcta de los datos de búsqueda. Sin embargo, si bien es cierto que el sistema predice correctamente la primera posición, no podemos decir lo mismo de las tres posiciones siguientes, donde las búsquedas nos exageran la importancia de los candidatos que quedaron tercero y cuarto respectivamente. Además, parece ser que el tema “*vote Le Pen*” fue menos activo en la semana previa a las elecciones, algo que no coincide con los resultados definitivos.

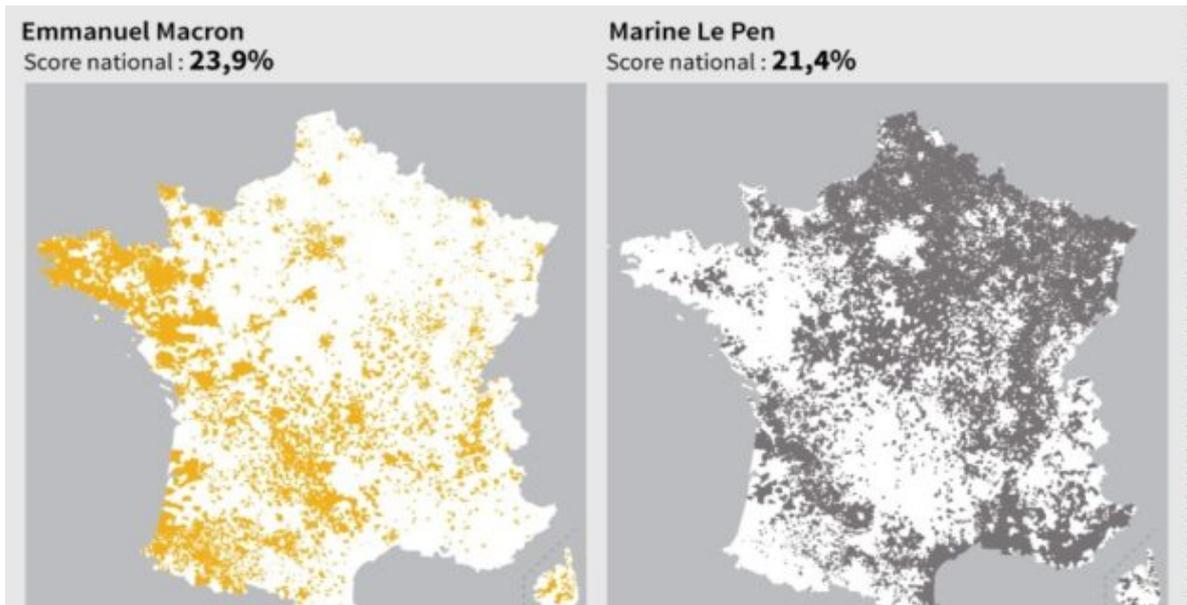
Analicemos ahora el interés por subregión de todos los candidatos a la primera vuelta de las elecciones (gráfica 22).



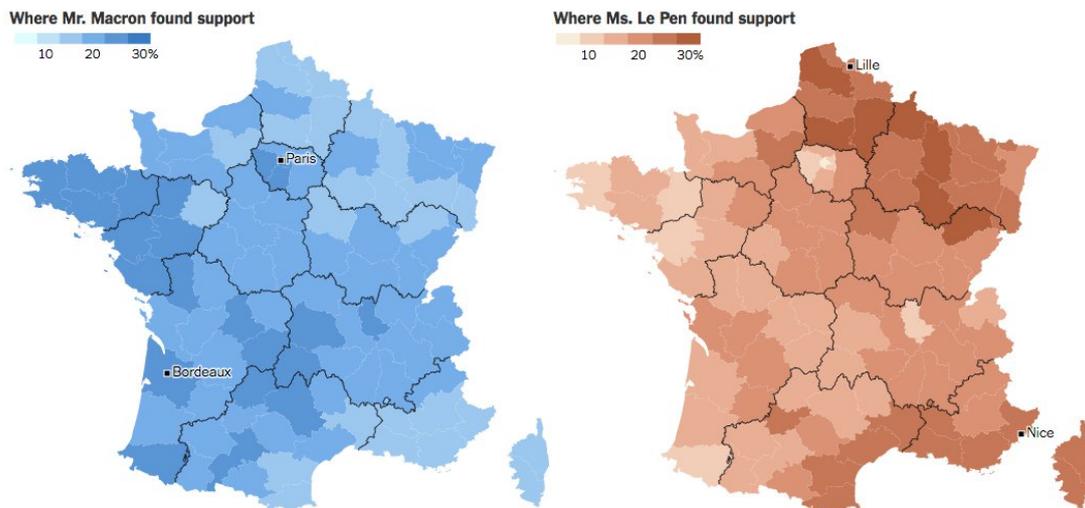
Fuente: Elaboración propia con los datos de Google Trends

En general, los datos reflejan un mayor interés por el candidato Emmanuel Macron en casi todas las provincias. Muy seguido, en algunas regiones, de François Fillon. Sin embargo, Marine Le Pen, que obtuvo finalmente el 21,4 % de los votos se ve infrarrepresentada en la gráfica 22.

Los dos candidatos resultantes de la primera vuelta fueron, por tanto, Marine Le Pen con el 21,4% de los votos, y Emmanuel Macron, con el 23,9% de los votos. Ambos candidatos quedaron bastante igualados a lo que resultados se refiere, sin embargo, obtuvieron votos de zonas significativamente distintas de Francia. Tal como reflejan los siguientes mapas elaborados por el New York Times:

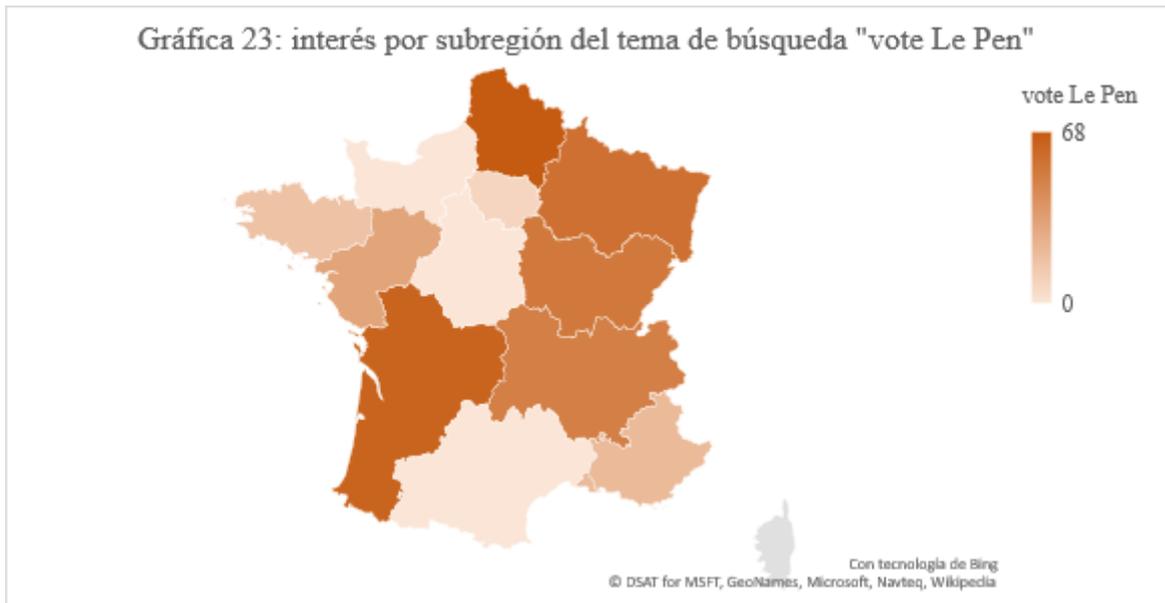


Fuente: New York Times

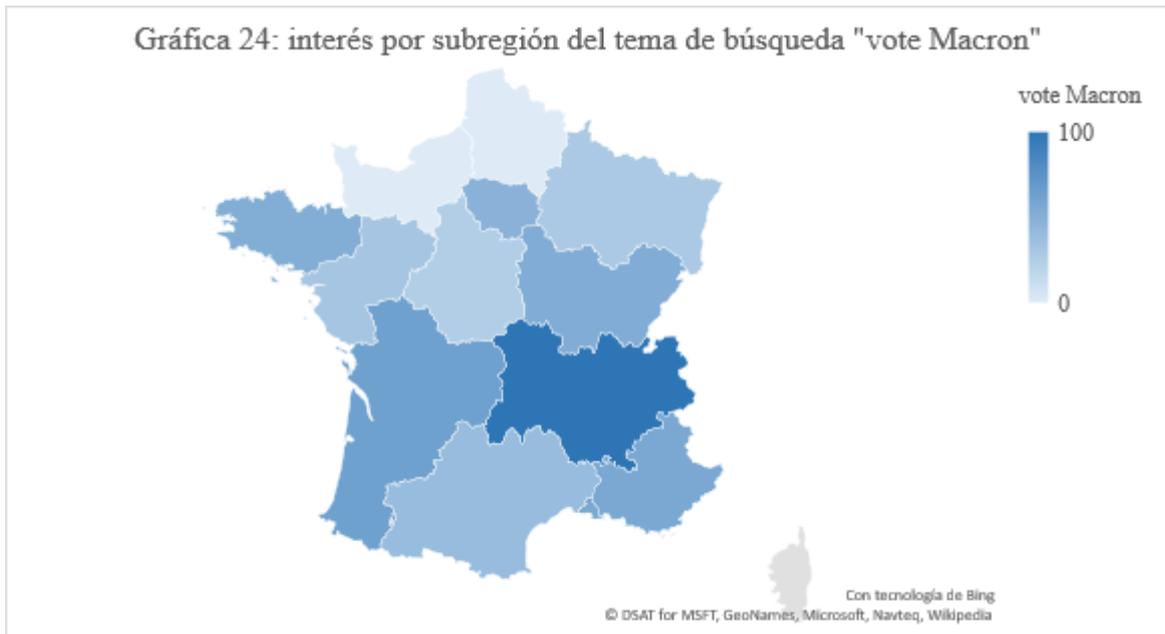


Fuente: New York Times

Los gráficos superiores evidencian un hecho anunciado ya por las encuestas. El norte desindustrializado de Francia, des de Dijon hasta Lille, se siente especialmente afectado por la crisis económica de los últimos años, y, en consecuencia, votaron en su mayoría, por la opción más radical, la de Marine Le Pen. En contraste, el sur y el oeste francés optaron por el candidato independiente Emmanuel Macron. Si cotejamos estos datos con los análisis de tendencias de Google Trends, observamos cómo estos resultados coinciden, casi por completo, con el interés por subregión que despertaron ambos candidatos en la red (graficas 23 y 24).



Fuente: Elaboración propia con los datos de Google Trends

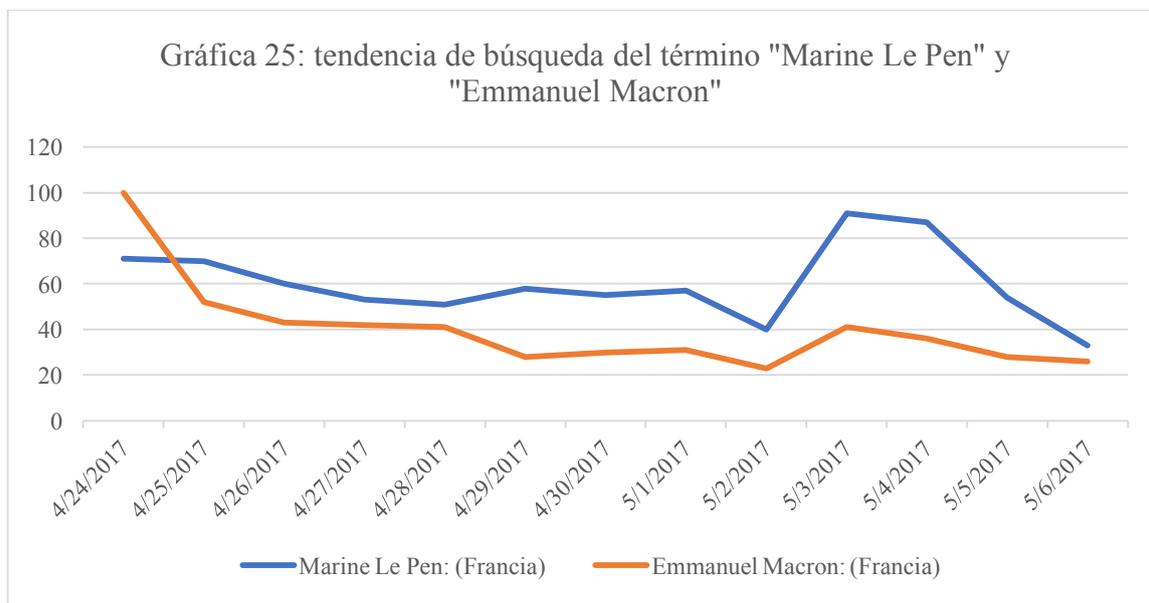


Fuente: Elaboración propia con los datos de Google Trends

Sin embargo, como se ha explicado anteriormente, estos datos no son realmente significativos para predecir el resultado final de las elecciones, pues corresponden solamente a la primera vuelta de las elecciones. No obstante, las gráficas superiores sí nos dan una idea de la importancia que tiene la red en las elecciones.

5.3.2 La segunda vuelta

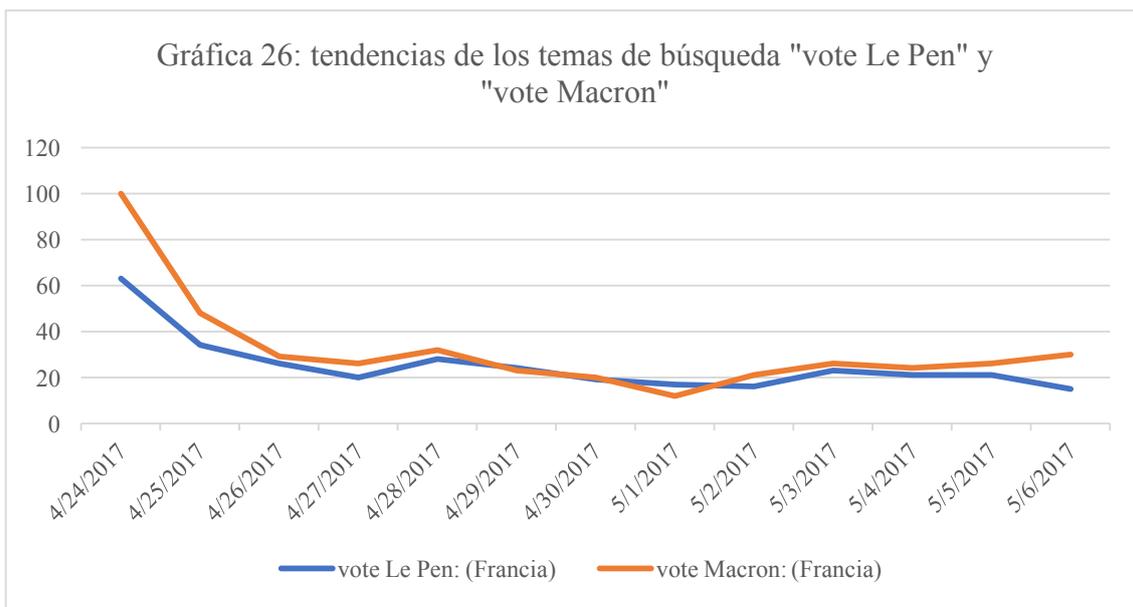
La segunda vuelta de las elecciones francesas se produjo el 7 de mayo del 2017. Eso nos deja un plazo de dos semanas exactas entre la primera vuelta y la segunda. Los siguientes análisis de datos, por tanto, corresponden a un lapso de tiempo relativamente reducido, des del 24 de abril del 2017 hasta el 6 de mayo del 2017. En la gráfica 25 observamos cual fue la tendencia de búsqueda de ambos candidatos durante ese periodo.



Fuente: Elaboración propia con los datos de Google Trends

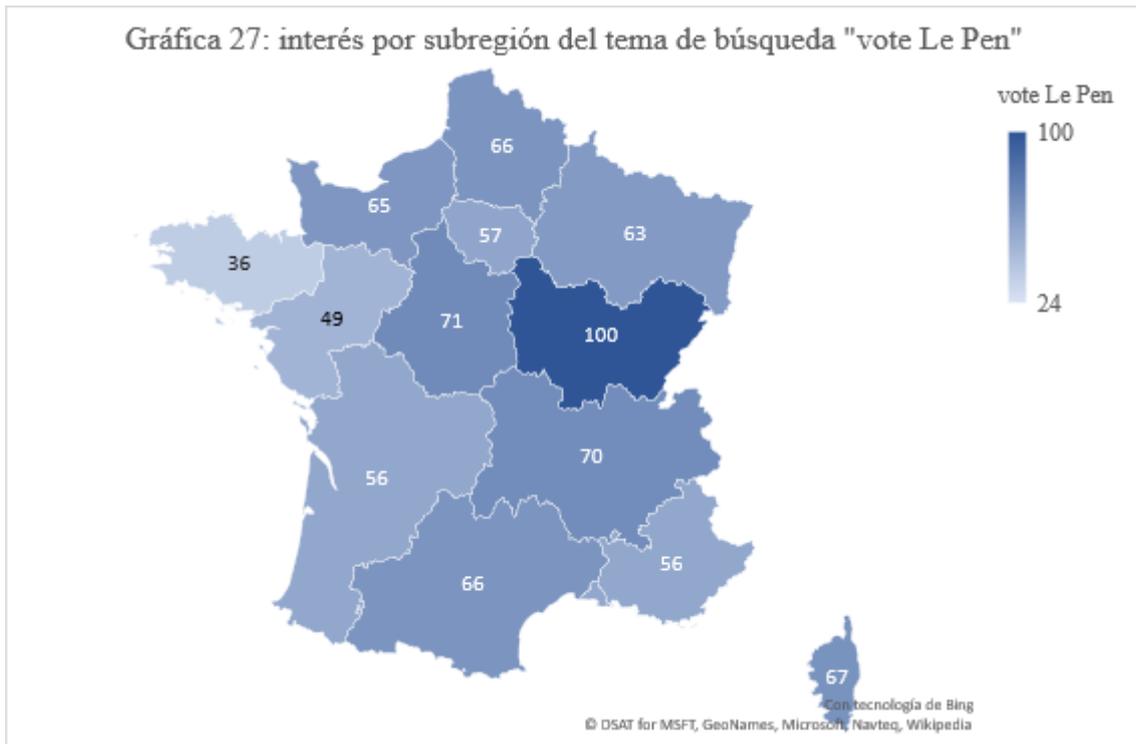
El 24 de abril, con la reciente victoria de Macron y Le Pen, ambos candidatos gozaban de cierta popularidad en la red, especialmente Macron, el candidato favorito a liderar Francia. Observamos, sin embargo, como el número de búsquedas de Macron desciende de forma brusca entre el 24 y el 25 de abril. Marine Le Pen, no obstante, sigue siendo un término bastante buscado durante toda la semana posterior, superando las consultas de su contrincante. El número de consultas a Marine Le Pen se dispara durante la noche del último debate electoral entre los dos candidatos (02/05/17) pero empieza a descender abruptamente los días anteriores a la segunda vuelta. Las tendencias de búsqueda de Macron se estabilizan sobre los 40 puntos durante todo el lapso de tiempo.

Debemos recordar que estos datos no son significativos para predecir los resultados, pues las consultas que se muestran en la gráfica 25 no están todavía filtrados. Es decir, corresponden a cualquier consulta en Google por la que se haya introducido el término "Marine Le Pen", ya sea esta buena o mala. Procedamos, por tanto, a analizar las tendencias de búsqueda de ambos candidatos, filtradas con la palabra "vote + nombre" (gráfica 26).

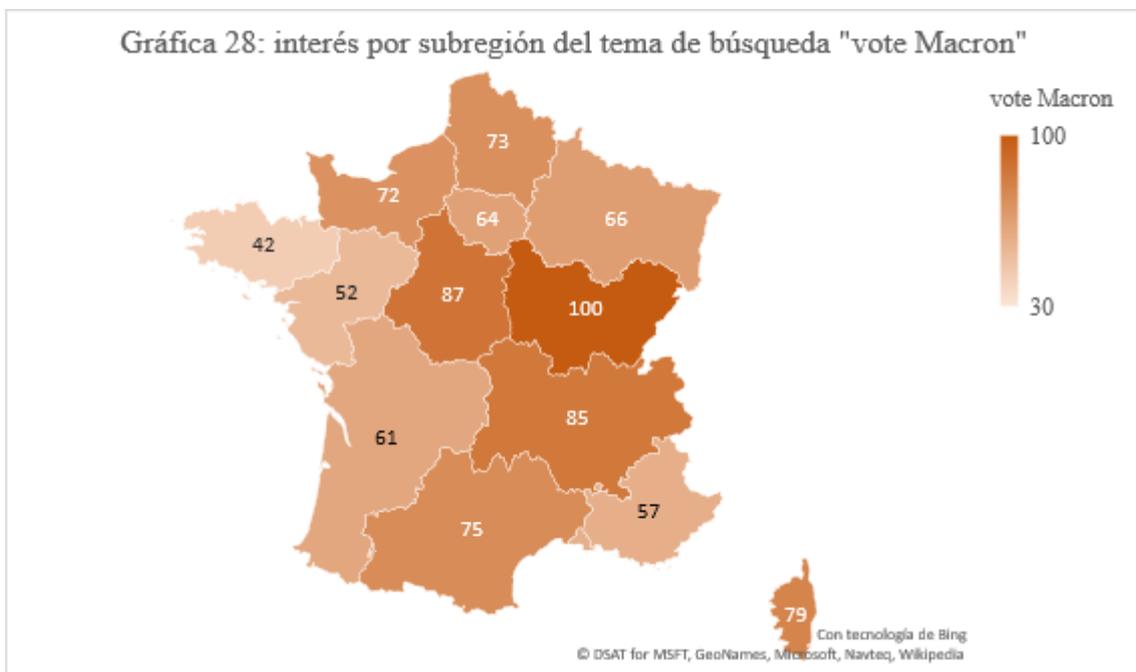


Fuente: Elaboración propia con los datos de Google Trends

Los resultados filtrados de la gráfica 26 nos ofrecen unos datos más realistas del interés que había por ambos candidatos. Si comparamos el gráfico superior con la gráfica 25, observaremos, en primer lugar, que el número de consultas por Marine Le Pen se ha reducido drásticamente. Los temas de búsqueda “*vote Le Pen*” y “*vote Macron*” se estabilizaron alrededor de los 20 puntos durante las dos semanas anteriores a la segunda vuelta. Sin embargo, en la semana posterior a las elecciones, el tema de búsqueda “*vote Macron*” parece superar en número a las búsquedas que incluían los términos “*vote Le Pen*”, un hecho que queda patente en el fin de semana electoral, cuando empiezan a diferenciarse claramente las tendencias. A lo que refiere el interés por subregión de ambas opciones, el mapa de Francia quedaría de la siguiente manera (gráficas 27 y 28):



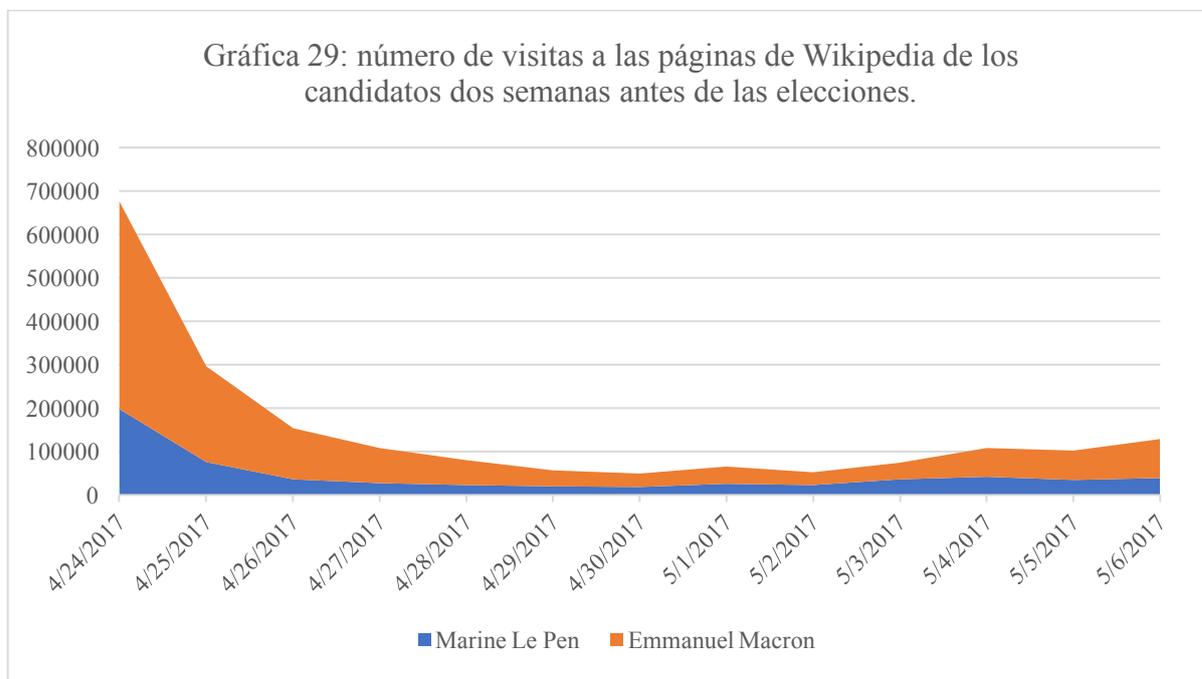
Fuente: Elaboración propia con los datos de Google Trends



Fuente: Elaboración propia con los datos de Google Trends

En general, vemos como las búsquedas por “*vote Le Pen*” fueron menores en todas las regiones de Francia. Si bien es cierto que la actividad del tema de búsqueda “*vote Le Pen*” fue mayor en el norte de Francia, en ningún momento parece superar a la actividad de búsqueda de su contrincante. Merece una mención especial, sin embargo, el caso de la

región de Borgoña, donde la actividad en la red fue alta para ambos candidatos y donde las dos opciones experimentaron el mismo aumento de búsquedas. En definitiva, las tendencias de Google nos indican claramente que la opción de “*vote Le Pen*” no es más popular en ninguna región de Francia. Comparemos ahora estos datos con el número de visitas de las páginas de Wikipedia de los candidatos (gráfica 29):

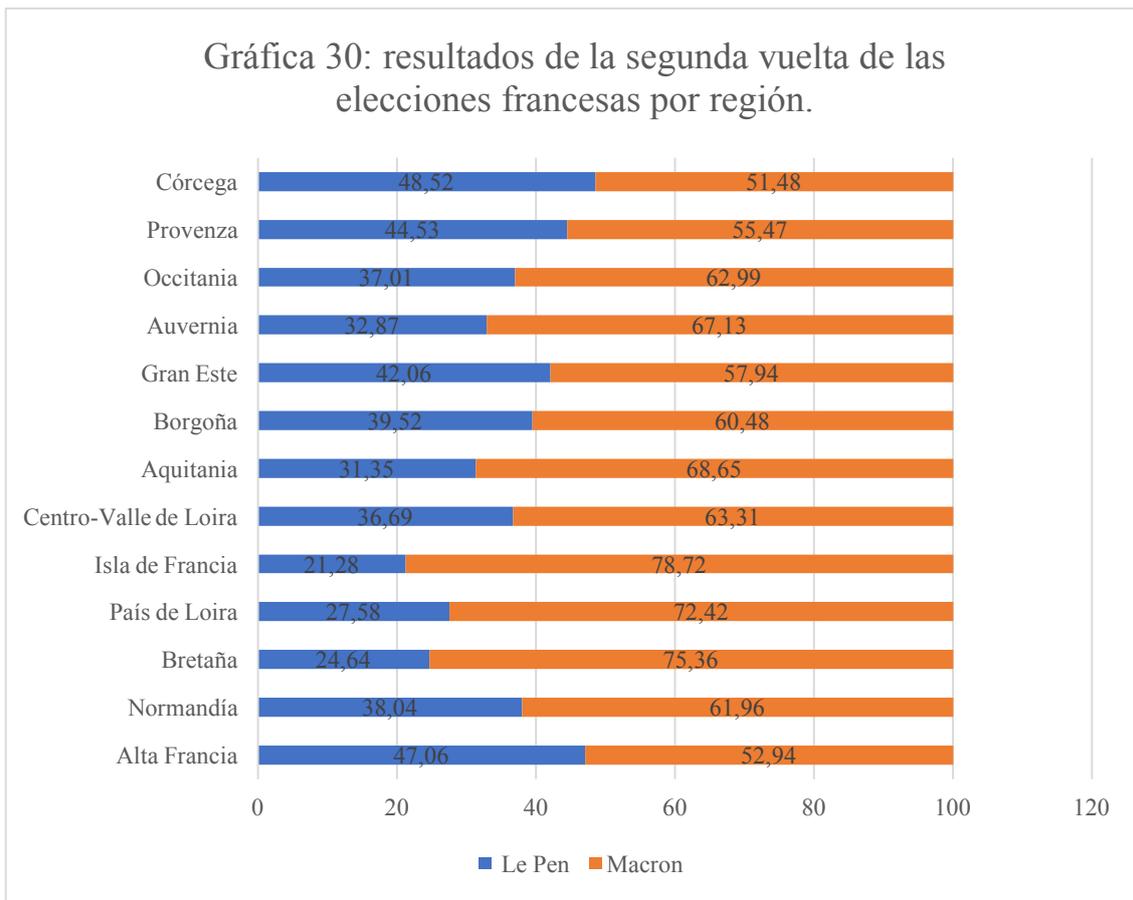


Fuente: Elaboración propia con los datos de Wikipedia page views counter

Durante el lapso de tiempo que transcurrió entre la primera y la segunda vuelta de las elecciones francesas, la página del candidato independiente, Emmanuel Macron, fue vista un total de 1.584.659 de veces en su versión francesa. Con un promedio de 121.897 visitas al día³⁶. Por su lado, la página de la candidata del Front National, Marine Le Pen, fue vista un total de 969.171 veces, rondando las 74.552 visitas al día. Como evidencia la gráfica 29, el número de visitas a la página de Emmanuel Macron superó a las de Marine Le Pen en todos los días de la campaña. Los datos de la Wikipedia, por tanto, coinciden con las tendencias observadas en Google Trends. No hubo un mayor interés por Marine Le Pen en las redes que el que hubo por el candidato Emmanuel Macron. Estos datos son alentadores, pues nos demuestran que si bien es cierto que hubo más publicaciones y búsquedas que incluían el nombre de Marine Le Pen (véase la gráfica 25), no gozó de más popularidad. Las tres últimas gráficas, las gráficas 27, 28 y 29, demuestran que los datos filtrados encajan, casi de manera perfecta, con los resultados de la segunda vuelta de las elecciones (gráfica 30).

³⁶ Datos disponibles en:

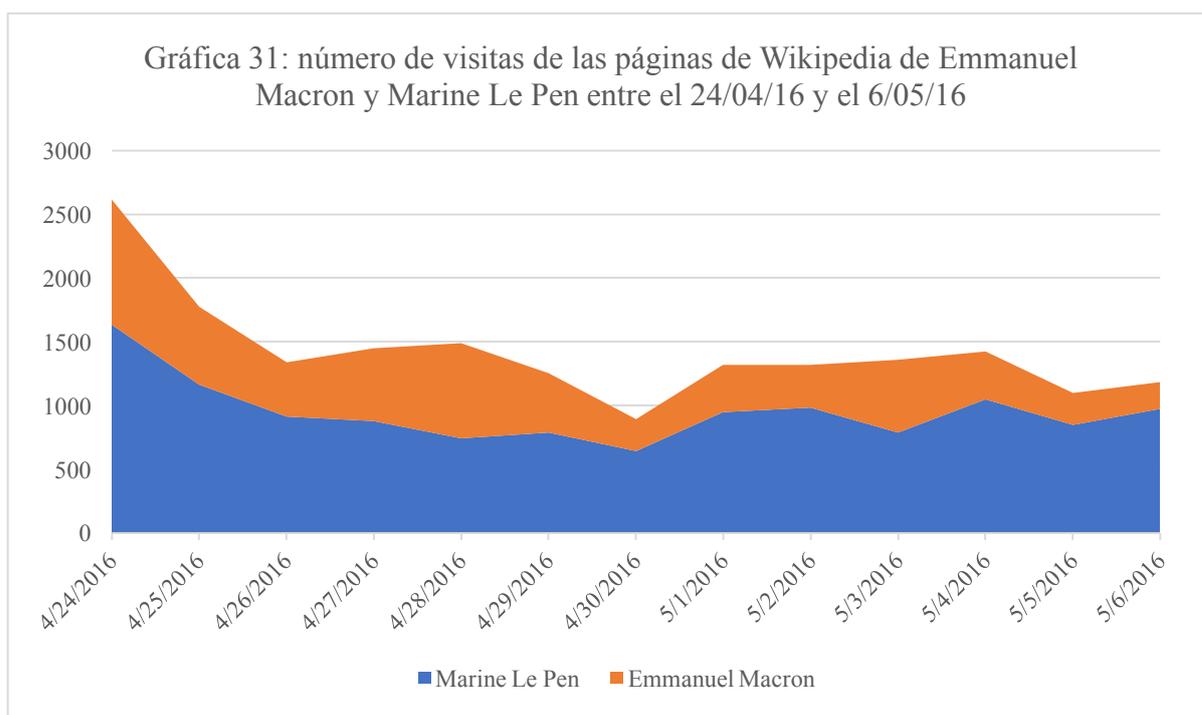
https://tools.wmflabs.org/langviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2017-04-24&end=2017-05-06&page=Emmanuel_Macron



Fuente: Elaboración propia con datos del Ministerio de Interior de Francia y NYT.

La gráfica 30 nos presenta los resultados definitivos de las elecciones francesas, extraídas del Ministerio Interior de Francia. Tal y como indicaban las tendencias de búsquedas de Google Trends, en ninguna región de Francia los resultados de Le Pen superaron a su homólogo liberal, Emmanuel Macron. Las zonas donde Marine Le Pen fue más activa en las búsquedas, Gran Este y Alta Francia, con 66 y 63 puntos respectivamente, son también las regiones donde más votos obtuvo la candidata del Frente Nacional. La única discrepancia entre las tendencias de Google y los resultados finales es el caso de Borgoña, donde, en lugar del empate que indicaba el buscador, se produjo una holgada victoria del candidato Emmanuel Macron. Los motivos que llevaron a esta diferencia no están del todo claros.

Nos queda, por último, comparar el tráfico de red de las páginas de Wikipedia de ambos candidatos para saber, si efectivamente y de acuerdo con los dos casos anteriores, se produjo un incremento del volumen de búsquedas y navegación. En la gráfica 31 se muestran el número de visitas de Emmanuel Macron y Marine Le Pen justamente un año antes de las elecciones.



Fuente: Elaboración propia con los datos de Wikipedia page views counter

La página de la candidata del Front National, Marine Le Pen, recibió un total de 15.609 visitas entre los días 24/04/16 y 6/05/16 en su versión francesa³⁷. Con un promedio de 1.201 visitas al día. Los datos son sorprendentes, si comparamos estas visitas con los datos de la gráfica 29 (969,171 visitas) observamos un aumento del 6.109%. El caso de Emmanuel Macron es parecido. Durante el mismo periodo del 2016, la página del candidato independiente alcanzó las 67.753 visitas³⁸. Su página era vista un total de 6.108 veces al día en su versión francesa. No obstante, estos datos parecen ínfimos si los comparamos al tráfico diario que recibió su página durante las elecciones y que queda reflejado en la gráfica 29. El aumento del periodo electoral corresponde a un 2.238%.

En definitiva, podemos afirmar sin ninguna duda que se produce un enorme aumento del tráfico hacia página como Wikipedia durante los periodos electorales. Esto es especialmente relevante para entender la forma en la que la población se informa antes de ir a votar.

³⁷ Disponible en: https://tools.wmflabs.org/langviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2016-04-24&end=2016-05-06&page=Marine_Le_Pen

³⁸ Disponible en: https://tools.wmflabs.org/langviews/?project=en.wikipedia.org&platform=all-access&agent=user&start=2016-04-24&end=2016-05-06&sort=views&direction=1&view=list&page=Emmanuel_Macron

6. Conclusiones:

A continuación, me gustaría enumerar brevemente las conclusiones extraídas de este estudio para, después, tratar cada afirmación en un contexto general. Las conclusiones extraídas del presente trabajo son siete, siendo estas:

1. Las predicciones electorales a través de medios de análisis de tendencias en internet como las presentadas en este texto han demostrado ser más fidedignas si se realizan a corto plazo, a saber, dos semanas anteriores al día de las elecciones. Y deben hacerse sobre hechos programados, no hipotéticos.
2. Los análisis presentados en este trabajo demuestran que a medida que se acerca la fecha de las elecciones, más precisas se vuelven las predicciones. En conclusión, el plazo ideal para realizar una proyección fiable a partir del Big Data es el de una semana antes de las elecciones.
3. Debido a la heterogeneidad de la información que aparece en un buscador y la imposibilidad de saber la intención real de un usuario cuando busca a un determinado candidato, es necesario filtrar la información antes de proceder a analizarla. No hacerlo, crearía una noción falsa de las tendencias, como demuestra la comparación entre las gráficas 18-21 y 25 -26.
4. El interés por subregión de servicios como Google Trends son un buen indicador de la tendencia de voto de una región geográfica. Sin embargo, estas indicaciones han demostrado ser muy genéricas y nada concretas. Con numerosas excepciones en los dos casos analizados. Por lo tanto, si bien ofrecen una tendencia general, no deberían ser tomadas como datos objetivos.
5. Los sucesos en el mundo real alteran de forma notable los datos de Google. De esta forma, observamos un enorme aumento en el tráfico de red durante acontecimientos trágicos, como la muerte de la diputada laborista Jo Cox el 16 de junio del 2016 (gráficas 4 y 5).
6. El contador de visitas de las páginas de Wikipedia es un método mucho más preciso para predecir tendencias de voto que Google Trends. Lo evidencian los tres casos prácticos realizados.
7. Un mayor aumento del tráfico de red hacia páginas como Wikipedia durante periodos electorales (gráficas 9, 10 y 31), nos demuestran lo importante que son estos servicios a la hora de ayudar a decidir al votante.

6.1 Reflexiones finales

La predicción de acontecimientos electorales a través de métodos de análisis de tendencias en internet es todavía un campo de estudio poco tratado. En este trabajo se ha intentado sondear una de las muchas aplicaciones posibles del estudio de tendencias, aplicándolo a un campo de estudio tan concreto como sugestivo, el de la demoscopia. Las herramientas utilizadas, como demuestran las conclusiones extraídas, son rudimentarias. Sin embargo, no podemos caer en el error de creer que eso las hace menos fiables. Google se ha convertido en pocos años en el buscador hegemónico de la red, y sus millones de usuarios lo hacen la base de datos más grande de la historia de la humanidad. En este trabajo, sin embargo, ha demostrado que su grandeza equivale a su heterogeneidad. Las tendencias observadas en Google Trends, como patenta el presente trabajo, quedan lejos de ser perfectas, y no sería correcto usarlas como fuente principal para ninguna información periodística de calidad si antes no las hemos filtrado. No obstante, no podemos suscribir cada una de las palabras anteriores cuando nos referimos a las tendencias en Wikipedia. Como evidencian los tres casos analizados en este texto, la Wikipedia y su índice de visitas son un indicador relativamente fiable para predecir resultados electorales. Esto no significa que el índice de visitas en Wikipedia sea capaz de predecir resultados electorales con exactitud, pero sí nos muestra los principales patrones de votos. Si nos ceñimos a las gráficas 19 y 29 del caso francés, observamos como en ambos casos, igual que en la gráfica 8, los resultados del análisis de tendencias en Wikipedia predicen con cierta exactitud los resultados finales. No en términos concretos, sino más bien en términos generales; es decir, quién será el candidato ganador. Por desgracia, no podemos extraer esa misma conclusión del caso británico, simplemente por ausencia de datos fiables. Nuestra única fuente son páginas genéricas sobre el referéndum, que, en sí, no suponen ningún indicador de la intención de voto de los usuarios. Sin embargo, el análisis de tendencias en Wikipedia del caso británico sí nos ha permitido descubrir uno de los hechos más significativos de este trabajo: el aumento del tráfico de red hacia páginas como Wikipedia durante periodos electorales. Considero de especial relevancia esta conclusión, pues nos ayuda a entender en qué grado las personas recurren a la red para informarse antes de ir a votar.

Antaño, la única fuente fiable para entender un proceso electoral de una forma mínimamente actualizada eran los medios de comunicación. El votante leía o escuchaba la prensa si quería informarse sobre los candidatos. Hoy ese paradigma es cosa del pasado. El recurso más accesible para las nuevas generaciones es, y será cada vez más, el internet. Buscar en internet hoy corresponde al abrir una enciclopedia, leer un periódico y preguntar a un profesor de antes. Toda la información se encuentra al alcance de un par de clics. Esto puede parecer trivial, pero no lo es si hablamos de procesos electorales. La salud de la democracia tal y como la conocemos se mide en el acceso a la información que tienen sus ciudadanos. Es imperativo, por ello, que entendamos cómo funcionan los flujos de información de nuestro tiempo. ¿A quién vota el ciudadano?, ¿Con qué información ha llegado a elegir su candidato? Y, sobre todo, ¿de dónde la ha extraído? Los análisis presentes en este texto nos demuestran que si hay un recurso público al que se recurre antes de ir a votar, ese es la Wikipedia. Herramientas como las utilizadas en

este trabajo son esenciales para entender la forma en la que funciona nuestra democracia. Como afirman Hal Varian y Hyunyoung Choi en un estudio para Google, la red nos ayuda a entender lo que ocurre, “predice el presente” (Choi & Varian, 2009, pág. 2)³⁹. En una época en la que la demoscopia se ha visto entredicha por métodos más rápidos y fiables, es necesario replantear su futuro. No nos encontramos, ni mucho menos, ante el fin de la demoscopia; como han querido vaticinar algunos antes de tiempo, pero sí podemos decir que es el fin de la demoscopia tal y como la conocemos hoy. Con nuevas herramientas, accesibles a cualquiera, se podrán realizar en un futuro predicciones mucho más precisas.

Existe, sin embargo, un contrapunto a esta información, casi convertido en problema filosófico. ¿Conocer estas herramientas y usarlas para predecir lo que sucederá no provocará que aquello que iba a suceder cambie? El problema en cuestión escapa de nuestras manos y seguramente requiera de análisis posteriores para ser probado. Sin embargo, debo confesar que, en la línea en la que apuntaban H. Choi y H. Varian, estoy de acuerdo con afirmar que las herramientas de predicción de Big Data nos ayudan a entender el presente, más que a predecir el futuro (Choi & Varian, 2009). A lo sumo, podríamos decir que nos ayudan a predecir un futuro a corto plazo. Este hecho se desprende también de los casos prácticos presentados en este estudio. Si observamos las tendencias de búsqueda, tanto en Google como en la Wikipedia, ambas plataformas se vuelven más fiables a medida que se acercan las elecciones. Los datos son más numerosos y más precisos en función de la proximidad a la fecha decisiva.

Además, los análisis realizados en este trabajo completan los de Catherine Lui en el 2011 y los del profesor T. Yasserli en el 2014. Sendos investigadores llegaron a conclusiones similares: todavía era temprano para utilizar la red como método predictivo, por el simple hecho de que gran parte de la población, y, por tanto, del censo, todavía no se encontraba en ella (Lui, Metaxas, & Mustafaraj, 2011) (Yasserli, 2014). No obstante, hemos visto en este trabajo un enorme progreso en cuanto refiere a la fiabilidad de los datos. De las tres predicciones realizadas, ninguna fracasó en predecir el resultado, siempre y cuando se hubiera filtrado previamente la información. Los datos aquí presentados, por tanto, invitan al optimismo y nos sugieren que, en un futuro no muy lejano, estas predicciones sean todavía más precisas. Será tarea de futuros investigadores verificar esta tesis.

Por último, debemos también referirnos al provecho que estas conclusiones pueden tener para el periodismo. El periodismo y la demoscopia siempre han sido dos ámbitos muy cercanos el uno al otro. No se entiende, de hecho, demoscopia sin periodismo. La prensa, en su función de vigilante del poder e informante del pueblo ha sido siempre la encargada de aproximar ambos mundos. Política y votantes, poder y pueblo. Pero la política del siglo XXI ha cambiado y la forma en la que la prensa observa y analiza los procesos electorales debería cambiar también. No hacerlo provoca una enorme desconexión entre

³⁹ El estudio en cuestión se titula “*Predicting the present with Google Trends*”, en clara referencia a la capacidad de pronosticar tendencias a corto plazo que tiene la herramienta del buscador.

el mundo real y aquello sobre lo que se informa. Como admitió el propio New York Times algunos días después de las elecciones de noviembre de 2016 en una carta a sus lectores⁴⁰. Los análisis predictivos a partir de Big Data pueden realizarse con herramientas de acceso público, tal y como demuestra este texto, y podrían servir para complementar la información periodística que se ofrece al lector.

La demoscopia, como tantos otros sectores, ha infravalorado enormemente el papel de las nuevas tecnologías en la vida de las personas. La población ya no vota de la misma manera en la que lo hacía hace veinte años; tampoco se informa de la misma manera, y, mucho menos, piensa de la misma manera. ¿Los sistemas de predicción siempre van a fallar a partir de ahora? Naturalmente no, así lo demuestran las muy acertadas encuestas francesas. Pero no debemos ser ilusos. Las probabilidades de fallar en las elecciones francesas del 2017 eran escasas. La demoscopia tradicional trabaja bien con márgenes amplios, pero se pierde en los resultados ajustados. El motivo es claro: no tiene la capacidad suficiente para saber cómo se informa el votante antes de ir a votar, solamente tiene su palabra.

Tanto la demoscopia como el periodismo de los próximos años tendrán que entender esta nueva realidad: la mitad de los hechos que hoy acontecen en la vida real, transcurren en la red y sus efectos se dejan ver tanto en el espacio virtual como en el espacio real⁴¹. En definitiva, considero que el análisis de tendencias en internet otorga al periodista, y, por ende, al lector, una visión más plural de los hechos. Y, en consecuencia, puede disminuir la brecha que eventos como el Brexit o la elección de Donald J. Trump han creado entre el periodismo de datos (demoscopia) y la sociedad.

⁴⁰ Redacción, New York Times, *To Our Readers, From the Publisher and Executive Editor*, New York Times (13 de noviembre de 2016) Disponible en: https://www.nytimes.com/2016/11/13/us/elections/to-our-readers-from-the-publisher-and-executive-editor.html?_r=0

⁴¹ Como demuestra el caso del asesinato de la diputada Jo Cox en las tendencias de búsqueda (gráficas 1, 4, 5, y 6)

7. Bibliografía

- Sociedad de la Información en España. (2016). *Estudio Anual de Redes Sociales*.
Fundación Telefónica. Barcelona: Editorial Ariel.
- Arribas, B. A., & Montanera, R. (2016). *Estudio Anual de Redes Sociales*. IAB Spain.
ADGLOW.
- Barbaro, M. (9 de Noviembre de 2016). How Did the Media — How Did We — Get
This Wrong? *The New York Times*.
- BBC. (19 de marzo de 2014). El robot que escribe noticias debuta en Los Ángeles.
BBC.
- Bloggin Zenith. (10 de Enero de 2013). El triunfo de Obama en Internet: caso de
estudio de las campañas de 2008 y 2012. *Bloggin Zenith*.
- Choi, H., & Varian, H. (2009). *Predicting the Present with Google Trends*. Google Inc.
- Contreras, M. (2016). La Inteligencia Artificial MogIA predijo la victoria de Trump...
y las de Obama. *20minutos*.
- Dörr, K. N., & Hollnbuchner, K. (Abril de 2016). Ethical Challenges of Algorithmic
Journalism. *Digital Journalism*, 1-16.
- El País. (24 de Junio de 2016). Resultados del referéndum del "Brexit". *El País*.
- García, J. (2013). Sistema electoral de Francia. *Revistilla.es*.
- Giménez, A. (10 de 11 de 2016). Trump o la aplastante victoria del #BigData. *Agencia
Nacional de Noticias*.
- Gonzalo, M. (8 de Agosto de 2013). "Los datos masivos (o big data) son el nuevo oro".
eldiario.es.
- Juez, B. (29 de Diciembre de 2016). Donald Trump, tuitero en jefe. *El Mundo*.
- Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (Julio de 2016). Intrusion of software
robots into journalism: The public's and journalists' perceptions of news written
by algorithms and human journalists. *Computers in Human Behavior*, 71, 291-
298.
- Kharpal, A. (28 de Octubre de 2016). Trump will win the election and is more popular
than Obama in 2008, AI system finds. *CNBC*.
- Lewis, S. C. (2015). Journalism In An Era Of Big Data. *Digital Journalism*, 3, 321-
330.

- Lui, C., Metaxas, P. T., & Mustafaraj, E. (2011). *On the predictability of the U.S. elections through*. Wellesley: Wellesley College Digital Scholarship and Archive.
- Mars, A. (11 de Enero de 2017). Obama se despide alertando de las amenazas a la democracia. *El País*.
- Peña, I. d. (23 de Octubre de 2016). Underwood tenía razón: el 'big data' marcará las elecciones en EEUU. *El Confidencial*.
- Redacción. (13 de noviembre de 2016). To Our Readers, From the Publisher and Executive Editor. *New York Times*.
- Rodríguez, C. (29 de 03 de 2017). La historia de desamor entre Reino Unido y la Unión Europea. *El Mundo*.
- Sánchez, C. (18 de Febrero de 2016). Cuando las musas visitan al robot: así escribe poesía un algoritmo . *eldiario.es*.
- Tubella, P. (17 de Junio de 2016). Muere la diputada laborista Jo Cox tras ser tiroteada y apuñalada. *El País*.
- Valderrama, M. (7 de Mayo de 2017). 'Au revoir' Hollande, el presidente más impopular de la V República. *El Mundo*.
- Yasseri, D. T. (2014). *Can electoral popularity be predicted using socially generated big data?* Oxford: Oxford Internet Institute, University of Oxford.

8. Apéndice