EXTRACCIÓN DE DATOS TWITTER/ CIS (definir CIS por otro lado)

De cara a combinar la información oficial del INE y los diferentes ministerios con información sobre la intención de voto de los electores, para intentar estudiar una posible predicción electoral futura, trabajamos, además de las fuentes oficiales, con datos de Twitter y del CIS.

Twitter:

La literatura consultada respecto a la utilización de twitter como herramienta de predicción electoral nos lleva a esperar un resultado no demasiado ajustado u optimista. Más bien al contrario.

Aun así, puesto que el tema está en vías de desarrollo y es objeto de numerosas investigaciones nacionales e internacionales y tesis doctorales actuales, decidimos incluir la variable redes sociales, en concreto twitter, en nuestro análisis de los resultados electorales de las elecciones generales del último año.

Entre la bibliografía consultada destaca:

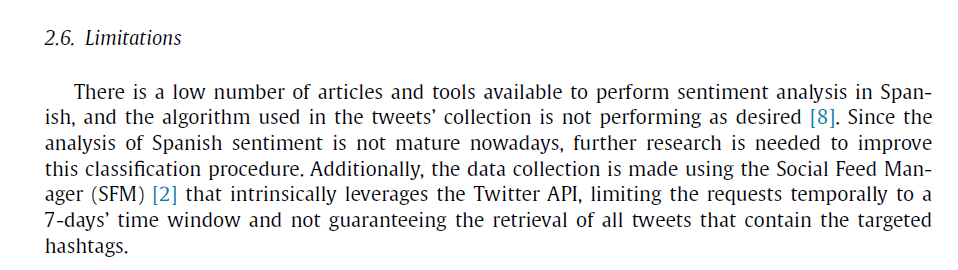
1.- sobre la importancia de tener en cuenta los bots en el análisis.

Este estudio trabaja con datos sobre las elecciones de noviembre de 2019 y lo usamos como referencia a la hora de decidir, entre otros, los hashtag a buscar y las fechas a considerar.

Twitter social bots: The 2019 Spanish general election data

<https://www.sciencedirect.com/science/article/pii/S2352340920309410?via%3Dihub>

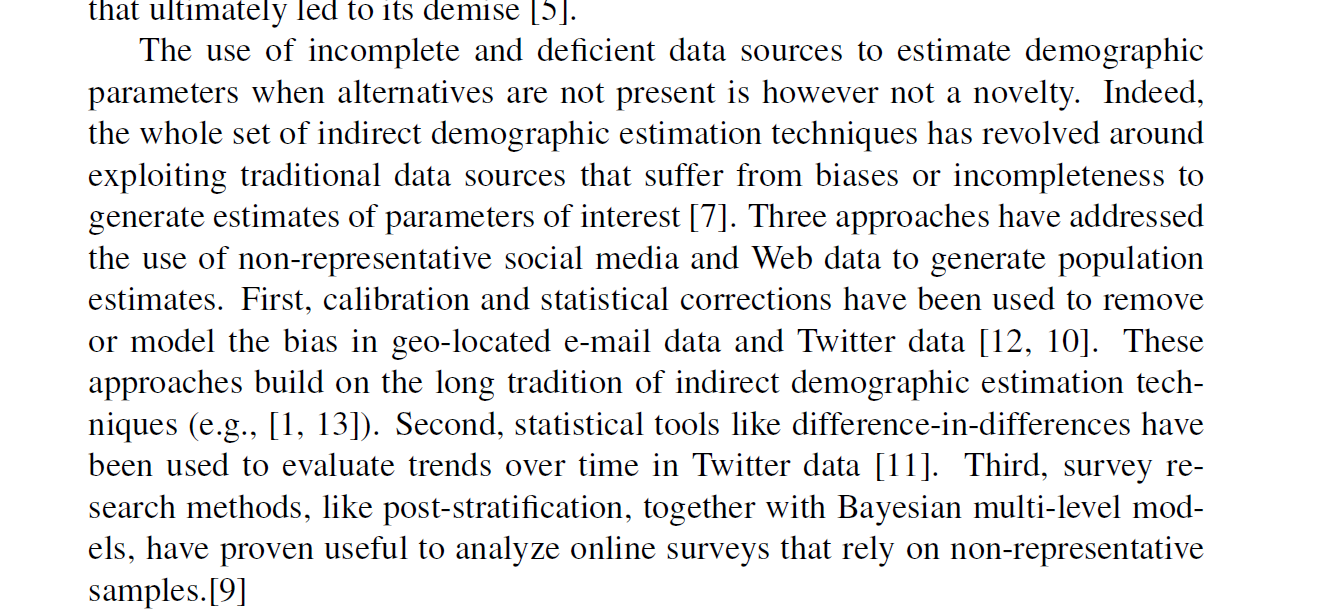
<https://github.com/CyberDataLab/botbusters-spanish-general-elections>



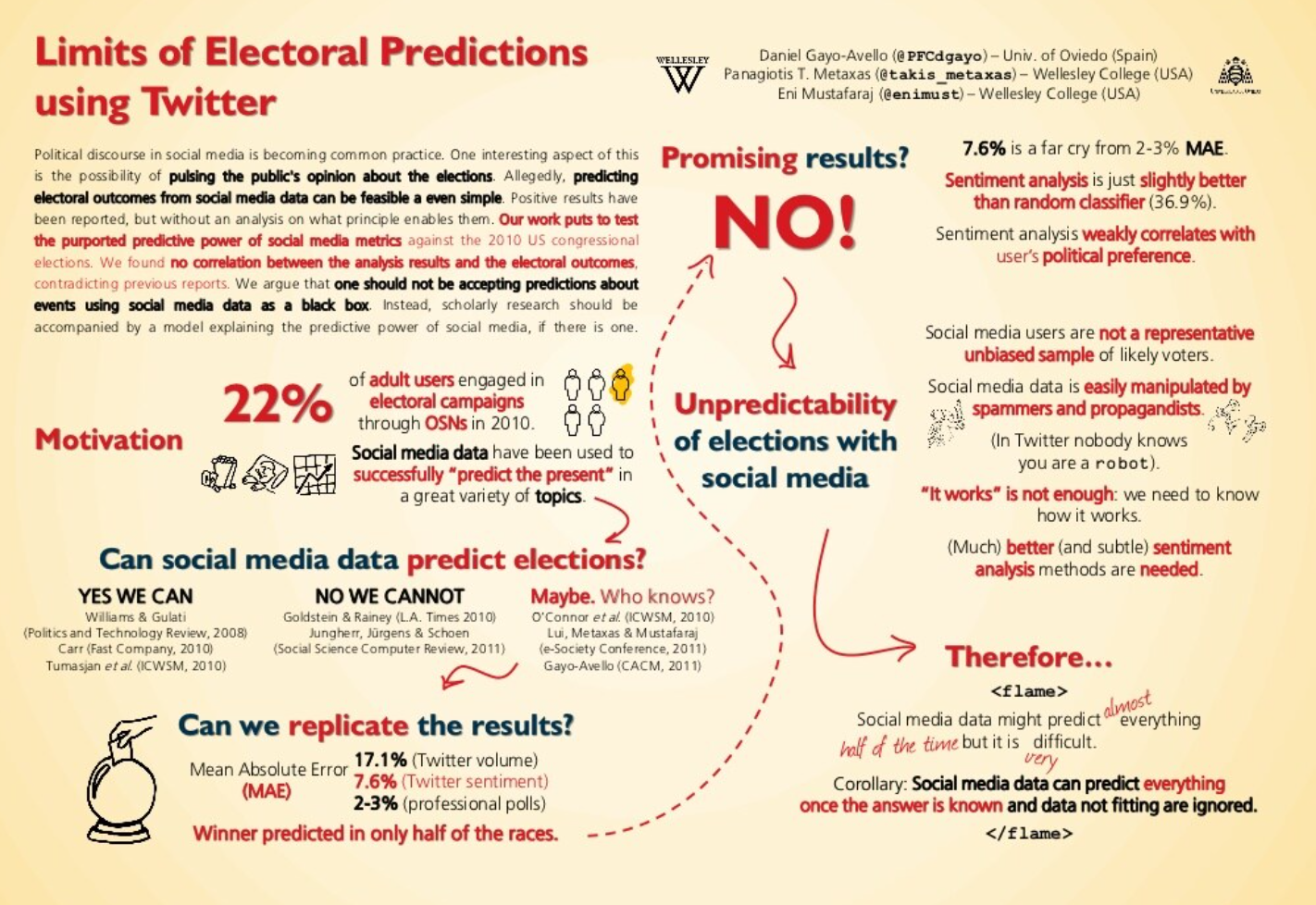
2.- Sobre la combinación de datos de encuestas tradicionales con análisis de intención de voto en redes sociales. Valoramos la opción de combinar los dos análisis con los datos oficiales aunque finalmente lo dejamos para una versión futura y realizamos análisis de uno y otro de manera independiente.

Combining Social Media Data and Traditional Surveys to Nowcast Migration Stocks

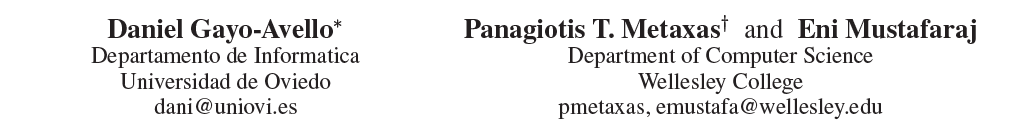
Emilio Zagheni, Kivan Polimis , Monica Alexander , Ingmar Weber , and Francesco C. Billari

[https://paa.confex.com/paa/2018/mediafile/ExtendedAbstract/Paper23641/Combining-social-media-and-ACS.pdf](https://paa.confex.com/paa/2018/mediafile/ExtendedAbstract/Paper23641/Combining-social-media-and-ACS.pdf)

3.- sobre los resultados a esperar de usar twitter para predecir resultados electorales.



Limits of Electoral Predictions Using Twitter.

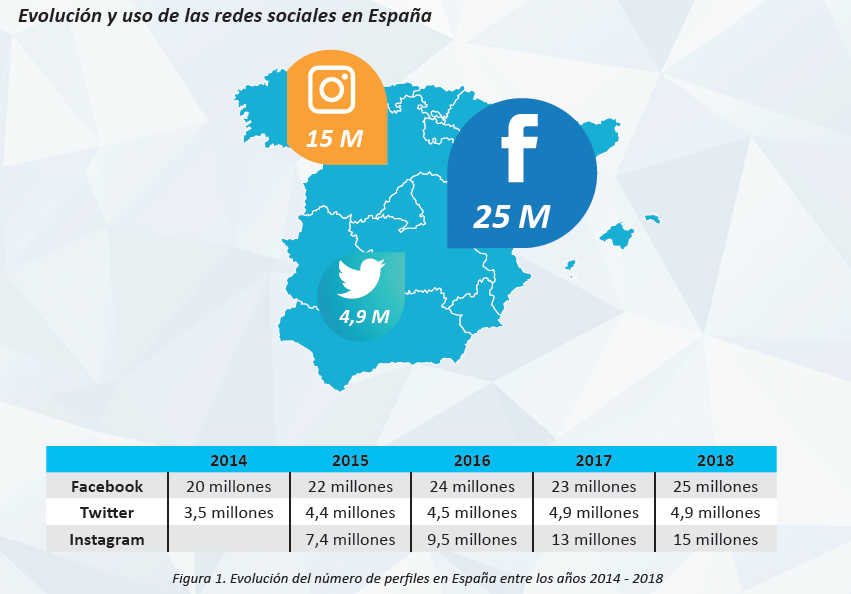


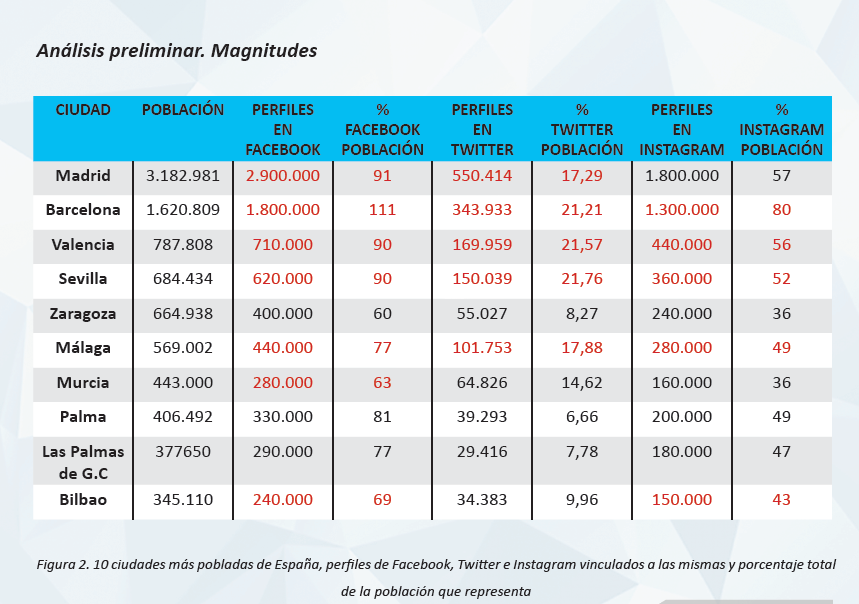
[Limits of Electoral Predictions Using Twitter (uniovi.es)](https://digibuo.uniovi.es/dspace/bitstream/handle/10651/11899/2862-14190-1-PB.pdf;sequence=2)

4.- Sobre los usuarios de redes sociales en España en 2019

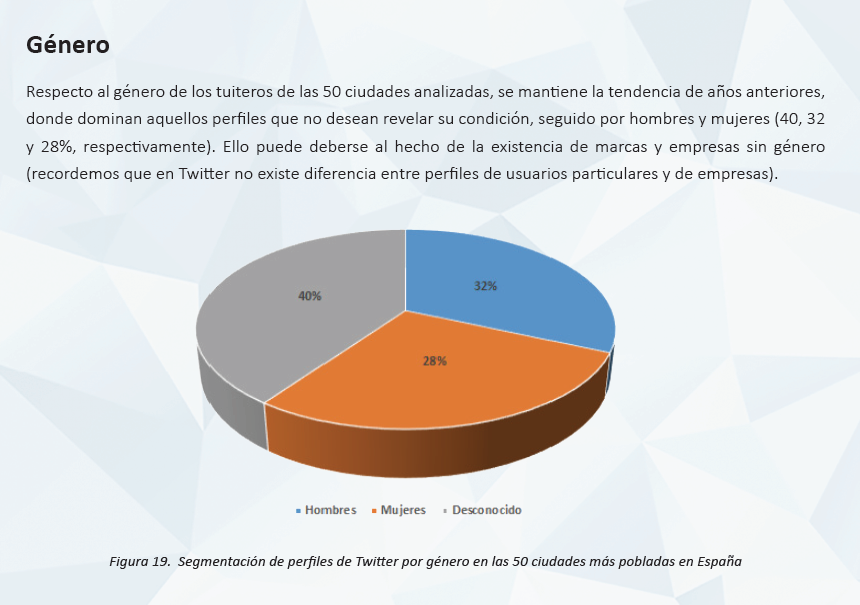
Para poder realizar un ajuste proporcional de los resultados de nuestra predicción a población a considerar dado el uso de twitter por los españoles.

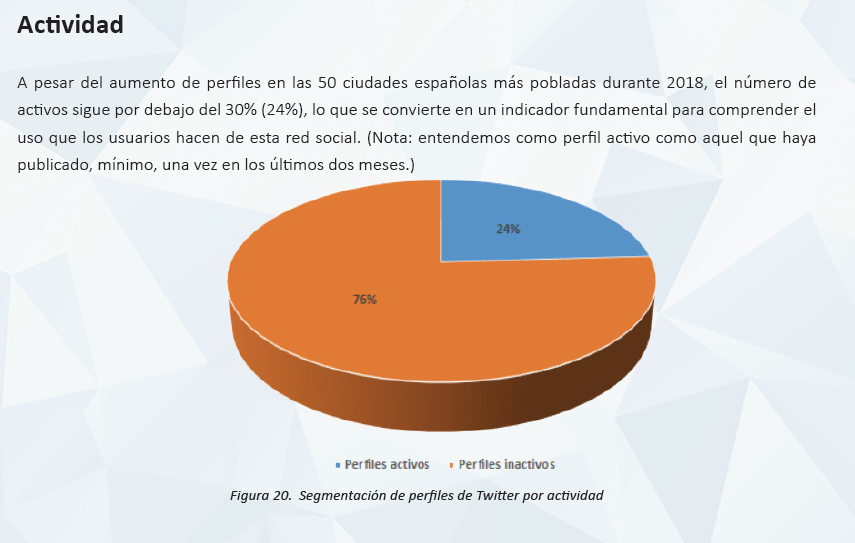
Estudio sobre los usuarios de Facebook, Twitter e Instagram en España. 2019

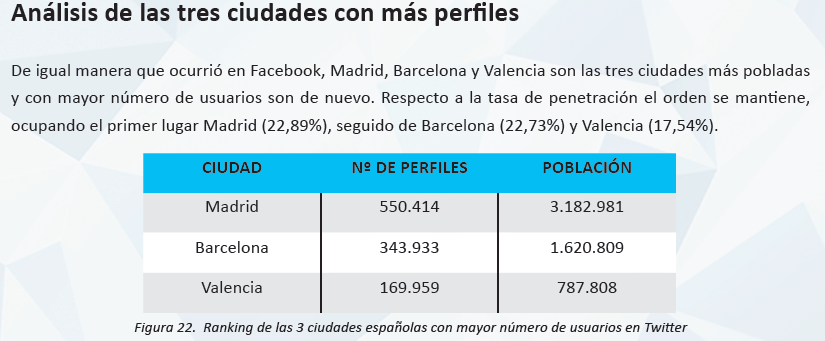












NUESTRO ANÁLISIS:

Aun con la evidencia en nuestra contra sobre la calidad de los resultados decidimos realizar nuestra propia búsqueda de tweets relacionados con las elecciones generales y los partidos para el año 2019 y valorar la posibilidad de realizar un análisis de sentimiento y detectar intención de voto.

EXTRACCIÓN DE DATOS:

Con la ayuda de la librería snscrape descargamos tweets tanto de las semanas anteriores a las elecciones de abril como de las de noviembre de 2019.

**Abril 2019**: 340210 Tweets detectados.

Hashtags buscados:

#28A, #ElDebateDecisivo, #ElDebateEnRTVE, #EleccionesGenerales28A, #Elecciones28A , #EleccionesL6, #Cs, #Ciudadanos, #VOX, #PSOE, #UnidasPodemos, #PartidoPopular, #PP, @vox\_es, @PSOE, @populares, @CiudadanosCs , @ahorapodemos

since:

2019-04-01 until:2019-04-29'

**Noviembre 2019**: Para noviembre realizamos dos búsquedas.

313913 Tweets detectados Hashtags electoral:

'10N OR EleccionesNoviembre2019 OR eleccionesgenerales10N OR Elecciones10N OR 10Noviembre OR 10NElecciones OR #Debate10N

324803 Tweets detectados hashtags partidos:

#Cs, #Ciudadanos, #VOX, #PSOE, #UnidasPodemos, #PartidoPopular, #PP, @vox\_es, @PSOE, @populares, @CiudadanosCs , @ahorapodemos

since:

2019-10-15 until:2019-11-11'

Finalmente analizamos a modo de ejemplo los datos de Abril de 2019.

Los notebooks están en el github numerados y los datasets globales en el bucket correspondiente de S3.

Twitter social bots: The 2019 Spanish general election data

Javier Pastor-Galindo a , ∗, Mattia Zago a , ∗, Pantaleone Nespoli a , Sergio López Bernal a , Alberto Huertas Celdrán b , Manuel Gil Pérez a , JoséA. Ruipérez-Valiente a , Gregorio Martínez Pérez a , Félix Gómez Mármol a

a *Department of Information Engineering and Communications, University of Murcia, Murcia, Spain* b *Telecommunication Software & Systems Group, Waterford Institute of Technology, Cork Rd, Waterford, Ireland*

formato para nombrarlo:

@article{Botbusters10N,

title = {Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election},

journal = {IEEE Transactions on Network and Service Management},

author = {Pastor-Galindo, J. and Zago, M. and Nespoli, P. and {Lopez~Bernal}, S. and {Huertas~Celdrán}, A. and {Gil~Pérez}, M. and Ruipérez-Valiente, J. A. and {Martínez~Pérez}, G. and {Gómez~Mármol}, F.},

year = {2020},

volume = {17},

number = {4},

pages = {2156-2170},

doi={10.1109/TNSM.2020.3031573}

}

Corpus ID: 15662762. **Limits of Electoral Predictions Using Twitter** @inproceedings{GayoAvello2011LimitsOE, title={**Limits of Electoral Predictions Using Twitter**}, author={Daniel Gayo-Avello and P. Metaxas and Eni Mustafaraj}, booktitle={ICWSM}, year={2011} }

1. **Gayo-Avello, Daniel** (2012). [¿Es Twitter un nuevo "Literary Digest"? No, No se pueden predecir elecciones con Twitter](http://www.catarata.org/libro/mostrar/id/806). In La comunicación política y las nuevas tecnologías (pp. 176-199). Los Libros de la Catarata.