Procesamiento del Lenguaje Natural, Revista nº 64, marzo de 2020, pp. 21-28 recibido 29-10-2019 revisado 09-01-2020 aceptado 10-01-2020

Detecting Influencers in Social Media using information from their followers

Detectando Influencers en Medios Sociales utilizando la información de sus seguidores

Javier Rodríguez-Vidal, Julio Gonzalo and Laura Plaza
Universidad Nacional de Educación a Distancia (UNED)
{jrodriguez,julio,lplaza}@lsi.uned.es

Abstract: Given the task of finding influencers of a given domain (i.e. banking) in a social network, in this paper we investigate (i) the importance of characterizing followers for the automatic detection of influencers; (ii) the most effective way to combine signals obtained from followers and from the main profiles for the automatic detection of influencers. In this work, we have modeled the discourse used in two domains, banking and automotive, as well as the language used by the influencers in such domains and by their followers, and used these Language Models to estimate the probability of being influencer. Our most remarkable finding is that influencers not only depend on their expertise on the domain but also on that of their followers, so that the more knowledge and number of experts among their followers, the more probability of being influencer a profile has.
Keywords: Learning to Rank, Web and social media search, Information extraction, Social Network Analysis, Natural Language Processing, Social Media Influencers

Resumen: Dada la tarea de encontrar influencers en un dominio dado (i.e. banking) en una red social, en este artículo investigamos (i) la importancia de caracterizar a los seguidores para la detección automática de influencers; (ii) la manera más efectiva de combinar señales obtenidas de los seguidores y de los perfiles principales para la detección automática de influencers. En este trabajo, hemos modelado el discurso usado por los usuarios en dos dominios, automotive y banking, así como el lenguaje utilizado por los influencers en dichos dominios y por sus seguidores, y utilizamos estos Modelos de Lenguaje para estimar la probabilidad de ser un influencer. Nuestro mayor descubrimiento es que los influencers no sólo dependen de su conocimiento sobre el dominio sino del de sus seguidores; por lo tanto, cuanto mayor conocimiento y número de expertos haya entre sus seguidores, mayor será la probabilidad que el perfil sea de un influencer.
Palabras clave: Learning to Rank, Búsqueda Web y en Medios Sociales, Extracción de Información, Análisis de Redes Sociales, Procesamiento de Lenguaje Natural, Influencers en Redes Sociales

1 Introduction

In traditional marketing it is imperative to know who is talking about an entity. Opinions of anonymous people do not have the same impact as opinions of special users, well-known people within communities, and who have the power to change the opinions of other users. These kind of users are known as *influencers* or *opinion-makers*.

Before the advent of Social Media, people with the capacity of influencing the public opinion in a given domain were few and easy to identify: journalists from mass media, authorities with academic degrees and proved expertise, politicians, media owners, celebrities, etc. In practice, editorial boards and lobbies could effectively decide what information and what opinions reached the masses, and how. Public Relations (PR) for organizations and individuals were, then, a matter of addressing a few opinion makers to shape their reputation, i.e., how their image was projected to the public opinion. Social Media has significantly complicated matters for organizations from the point of view of Public Relations. Monitoring and managing social

ISSN 1135-5948. DOI 10.26342/2020-64-2

media brings unprecedented opportunities to know and interact with clients and stakeholders, but it renders previous PR methodologies obsolete. One of the key aspects of Online Media, and of Social Media in particular, is that any citizen is a candidate to become influential: it is no longer possible to narrow the filter to media owners, journalists, academic experts and other standard profiles. In this context, one of the key aspects of Online Reputation Monitoring (ORM) is to detect which social media profiles have the capacity of influencing the public opinion and, therefore, creating opinion and shaping the reputation of organizations, companies, brands and individuals (Madden y Smith, 2010).

Just as there are profiles of influencers, there are also other kind of users in Social Networks that support them and serve as a loudspeaker for the propagation of the ideas of influencers, they are called *followers*. They are in charge of spreading the ideas of an influencer, either by retweeting a post or by generating new texts from influencers' ideas.

For this reason, in this work we want to check if we can detect influencers using information from followers. In particular, we want to know if the language used by followers is an important factor for establishing whether or not a user is an *influencer*. We distinguish here two different characteristics that define influencers: (i) the possession of knowledge about the *domain* of study (i.e. brokers in economy, mechanics in automotive, etc.) or (ii) the capacity of persuade other people (*authority*) because they are well-known for the general public, for example celebrities, sportsmen, etc.

To verify our hypothesis and methods, we used data extracted from Twitter. There are many important reasons, from the point of view of ORM, to use this social network rather than others, to name a few: (i) it is *immediate*, breaking news appear and propagate first than in other social network; (ii) it is *global*, accessible in the whole world and (iii) it is *asymmetric*, there is no need to have the consent to follow an account.

2 Related Work

Influencers are a special kind of users in Social Networks. They are trustworthy to the members of their communities and their ideas are capable to change other people's mind about an entity, even jeopardizing the entity's reputation. Aral y Walker (2012) defend that, in order to predict the propagation of actions, it is important to use jointly the influence, the susceptibility and the likelihood of spontaneous adoption in the local network around individuals. But, as the authors point out, it is not clear whether influence and susceptibility are general signals or depend on the domain.

Detecting *influencers* in ORM has two distinguishing signals: first, the number of influencers is orders of magnitude lower than the number of non-influencers. Second, potential influencers are usually scanned by reputation experts, which use automatic filters as a preliminary step. Both signals are characteristic of search problems, where ranking is the most natural way of presenting results to the users (in this case, the reputation experts). This approach is followed by the RepLab campaign (CLEF, 2014), an evaluation forum for ORM that, in its 2014 edition, included the *author ranking* task. This task aimed to distinguishing the users with the most reputational influence from the less influential users. In our work, we address the author ranking task and use for our experiments an extension of the RepLab 2014 dataset, which is described in Section 3.3.1. In the following lines, we describe systems that participated in the competition.

The best system in the competition was (AleAhmad et al., 2014), which implemented the idea that people who are opinion makers will talk more about hot topics. Another approach proposed in the RepLab competition (Cossu et al., 2014) assumed that influencers tend to produce more opinionated content in tweets. (Vilares et al., 2014) used the confidence provided by the LibLinear classifier to rank the users according to their level of influence, higher confidence means higher influence. (Villatoro-Tello et al., 2014) used techniques for signal extraction and collected the most representative signals from each user's activity domain. The last participant (Lomena y Ostenero, 2014) used a small set of features based on the information that can be found in the text of tweets: POS tags, number of hashtags or number of links.

Subsequent to this competition, new studies have appeared that have worked on this dataset. Cossu, Labatut, y Dugué (2016) tested different signals and concluded that users from particular domains behave and write in their own specific way and using only text-based signals is enough to detect domain influencers. Nebot et al. (2018) used embeddings to represent each document. Rodríguez-Vidal et al. (2019) used Language Models to compute signals that model the degree of authority and domain knowledge of the profiles.

None of these previous studies have exploited the information related to the followers, more than using the number of followers signal provided by the Social Networks. In our work, we model the language that followers of a given profile use to communicate their ideas, to estimate their probability of being an authority and domain expert. Our hypothesis in this work is that the probability of being influencer grows with the number of experts who follow the profiles.

3 Methods

In this section we introduce the signals and the algorithms employed for the automatic detection of influencers in Twitter by analysing their followers.

3.1 Signals

Our research is based on the work of (Rodríguez-Vidal et al., 2019), who used the textual content of the user's posts as a signal for predicting whether or not the user is an influencers. The authors show that the text in the user's posts gives useful domain information (active users in the banking domain, for instance, will use the distinctive vocabulary of the banking domain), and also provides evidence for authority, the reason being that authorities have distinctive commonalities in the way they express their opinions or transmit information. Our hypothesis is that the language employed by the followers could also be an important factor to take into account to locate influencers. For this reason, we replicate the calculus of the Language Models used by (Rodríguez-Vidal et al., 2019), which obtain a probability distribution of words, $p'(w)$, in which words likely to be included in an author message in the domain of authors are assigned high probability values; whereas other words, including those that are very ambiguous or not domain-specific but occur in the domain of authors, receive marginalized values. This distribution of words $p'(w)$ is optimized using an Expectation Maximization procedure, in the r-th iteration, is defined as:

$$p'^{(r)}(w) = \frac{p(w|L,D) * Z(w)}{(\sum_{w' \in V} p(w'|L,D) Z(w'))} \tag{1}$$

where V is the vocabulary $w_1, ..., w_{|V|}$; L and D being the background and the target domain, respectively; $Z(w)$ is the Expectation-Step and is defined as:

$$Z(w) = \frac{(1-\lambda)p'^{(r-1)}(w)}{((1-\lambda)p'^{(r-1)}(w) + \lambda p(w))} \tag{2}$$

$p(w|L)$ and $p(w|D)$ are defined as follows:

$$p(w|L) = \frac{tf(w,L)}{\sum_{w' \in L}(tf(w'))} \tag{3}$$

$$p(w|D) = \frac{tf(w,D)}{\sum_{w' \in D}(tf(w'))} \tag{4}$$

The probability of an author a belonging to the language model D is finally computed as:

$$p(D|a) = \sum_{w}(p(D|w) * p(w|a)) \tag{5}$$

where

$$p(D|w) = Z(w)$$

$$p(w|a) \propto tf(w, Y)$$

being Y the set of tweets of the author a. After that, we compare the language of each follower with the language models of authorities (authority model) or with the language models of tweets belonging to the domain (domain model). Next, we extract from them some signals in order to explore the role played by followers in the detection of influencers. One of our main goals is to compare the utility of these signals with those extracted from the main's profiles in the work of (Rodríguez-Vidal et al., 2019).

3.1.1 Authority signals

These signals are extracted by comparing the discourse of the authorities modeled as explained in (Rodríguez-Vidal et al., 2019) with the language used by the each follower to determine the connections made by the main profile. The name, description and formula of these signals, are described below:

1. **Auth:** probability of being an authority. In order to compute this signal, we first obtain the language model for the set of followers and the language model of the authorities. We denote this signal as $P_{auth}(f)$, where $f \in F$ and F is the set of followers of a given profile.

2. **Not_Auth:** probability of being non-authority. The signal value is computed by comparing the language models for the set of followers and the language model of the non-authorities. Note that, due to the fact that the language used by the authorities and the non-authorities may contain common words, the probability of being authority is not the complementary of the non-authority probability. We denote this signal as $P_{\neg auth}(f)$, where $f \in F$ and F is the set of followers of a given profile.

3. **#_Foll_Auth:** number of followers being authorities. This signal indicates the quality of the connections made for the main profiles. It is expected that connections with the right people lead to a high probability of being influencer. In order to compute this signal, we have to estimate the probabilities of being authority and non-authority for each follower. We count the followers as influencers if they fulfil the following condition: $P_{auth}(f) - P_{\neg auth}(f) > 0$, where $f \in F$ and F is the set of followers of a given profile.

4. **Mod_Foll_Auth:** similar to the previous one, it computes if a main profile is well connected. In other words, if a main profile is followed by a high number of influencers. Like the previous signal, we have to compute, previously, the probability of being authority and not being authority. The final signal is calculated as: $\sum_{f \in F} P_{auth}(f) - P_{\neg auth}(f)$, where F is the set of followers of a given profile.

5. **Avg_Mod_Foll_Auth:** it calculates, on average, the degree of authority of the followers of each main profile. This signal is computed as $\frac{Mod\_Foll\_Auth}{num\_foll}$, where *num_foll* is the number of followers of the main profile.

6. **Prop_Foll_Auth:** ratio of followers being authorities. This signal is computed as $\frac{\#\_Foll\_Auth}{num\_foll}$, where *num_foll* is the number of followers of the main profile.

7. **Avg_Prob_Auth:** is the ratio of followers' influence. Higher values indicate that the main profile messages are validated and disseminated by several experts. This signal is calculated as: $\frac{P_{auth}(f)}{num\_foll}$, where *num_foll* is the number of followers of the main profile and f are the followers of a given profile.

8. **Sum_Foll_Auth:** is the sum of the probabilities of the followers of a main profile being influencers. This signal is calculated as $\sum_{f \in F} P_{auth}(f)$ where $f \in F$ and F is the set of followers of a given profile.

3.1.2 Domain signals

These signals are extracted by comparing the texts published in each domain (automotive and banking) with the language used by each follower to know if they possess some background knowledge about the domains. The name, description and formula of these signals, are described below:

1. **Dom:** it measures how well the discourse of the followers fits in a domain. In order to compute this signal, we first model the language of the follower set and the domains. We denote this signal as $P_{dom}(f)$, where $f \in F$ and F is the set of followers of a given profile.

2. **#_Foll_Dom:** number of followers, of a main profile, that belong to a domain. A follower f fits in a domain if she fulfils the following requirement: $P_{dom}(f) - P_{\neg dom}(f) > 0$. Note that $P_{\neg dom}(f)$ is the probability of not belonging to the domain. Due to that words can belong to different domains, the probability of belonging to a domain may not be the complementary to the probability of not belonging to it.

3. **Mod_Foll_Dom:** it computes the connections of a main profile inside a domain, in other words, whether or not a main profile is followed by people with some knowledge about a domain.

It is calculated as: $\sum_{f \in F} P_{dom}(f) - P_{\neg dom}(f)$, where $f \in F$ and F is the set of followers of a given profile.

4. **Avg_Mod_Foll_Dom:** it calculates, on average, the knowledge about a domain that the followers of a main profile have in other words, whether or not a profile is followed by experts in a domain. This signal is computed as $\frac{Mod\_Foll\_Dom}{num\_foll}$, where *num_foll* is the number of followers of given profile.

5. **Prop_Foll_Dom:** is the ratio of followers which belong to a certain domain. This signal is computed as $\frac{\#\_Foll\_Dom}{num\_foll}$, where *num_foll* is the number of followers of the main profile.

6. **Avg_Prob_Dom:** is the ratio of followers which belong to a certain domain. Higher values indicate that the posts published by the main profile can be viewed and confirmed by domain experts. This signal is calculated as $\frac{P_{dom}(f)}{num\_foll}$, where *num_foll* is the number of followers of the main profile and f are the followers of a given profile.

7. **Sum_Foll_Dom:** is the sum of the probabilities of the followers of a main profile of belonging to a domain. This signal is calculated as $\sum_{f \in F} P_{dom}(f)$ where $f \in F$ and F is the set of followers of a given profile.

3.2 Algorithms

The detection and characterization of influencers is covered as a ranking problem because it is the most natural way of presenting results to the reputation experts (RepLab 2014). We have compared the two approaches that obtained the best results for the identification of influencers in (Rodríguez-Vidal et al., 2019) to generate a ranking of users' profiles:

- **Direct Signal Rank Strategy (DSR):** each extracted signal (see section 3.1) generates a ranking of users. For instance, we can rank users by the number of followers being authorities that main profiles have. When we use two or more signals to produce a single rank, we apply a Borda voting step (Saari, 1999) to combine the ranks produced by each individual signal. If we have n elements to rank, the Borda voting lies in an ordination of the elements to consider for each signal individually in descending order assigning the higher value, in our case n, to the first element of the ranking, the $n-1$ value to the second element and so on. The combined ranking is produced by adding the values assigned to each element by every rank, and using this number to produce the final ranking.

- **Learning to Rank Strategy (L2R):** since we have training and test data, we take advantage of them and use a *machine learning* algorithm called *Learning to Rank* (Liu, 2009), to make more accurate rankings. The models created try to optimize the selected metric on the training data (in our case Mean Average Precision (MAP)). We have used MART (*Multiple Additive Regression Trees*)(Friedman, 2001) to generate rankings.

  Each main profile that L2R receives is represented as a 1-hot vector, whose length is the total number of followers that exist for all main profiles without repetition. Those positions corresponding to a real follower of a profile, is filled with its respective value (probability of being authority or belonging to the domain), the other positions contain 0 as a value, which indicates that is not a follower of the profile. To combine these vectors, we only concatenate them.

  This experimentation was carried out using the RankLib tool (Dang, 2012).

3.3 Experimental framework

One of our primary objectives is to determine how our method, which is based on signals extracted from posts published by the followers of a Twitter user, behaves for identifying influencers, and compare it with other approaches that only use information from the main profile. To do so, our experiments are performed using as main profiles the ones in the RepLab 2014 dataset. As we mentioned in the previous section, we select some signals from the followers' posts to estimate whether a user is an influencer or not.

3.3.1 Dataset

We follow the guidelines of the RepLab 2014 competition and use the Author Ranking dataset in our experiments (Amigó et al., 2014). In this task, systems are expected to "find out which authors have more reputational influence" for a given domain (automotive and banking). The required systems' output is a ranking of Twitter profiles according to their probability of being influencers with respect to the domain. The RepLab 2014 dataset consists of 7,662 Twitter profiles (all with at least 1,000 followers) related to automotive and banking domains. The profiles are divided in: 2,502 training profiles and 4,862 test profiles. Each profile contains: (i) author name; (ii) profile URL and (iii) 600 tweets published by the author (in English and Spanish). Reputational experts manually assessed each profile as influencer or not.

Since the RepLab 2014 Dataset does not supply information related to the followers of its profiles (beyond the number of followers they have), we had to collect the names and tweets of the followers of each profile in the RepLab dataset. The followers retrieved are those whose profiles were created by the extraction time of RepLab (1st June 2012-31st December 2012) and have some post published during that period of time. We gathered 600 tweets per follower, but due to the time elapsed since the dataset was built, some tweets have been lost causing some of the followers to have less than 600 tweets. Despite the main profiles were manually assessed (as influencers or not influencers) by reputation experts, we do not have such information for the followers' profiles.

This extraction was carried out using GetOldTweets-java tool (Jefferson-Henrique, 2016).

3.3.2 Metrics

Mean Average Precision (MAP) (Manning, Raghavan, y Schütze, 2008) is the official metric for RepLab 2014 competition so that, in order to compare us with the state-of-the-art systems, we have also used it. This metric measures the average precision obtained for the top k documents after each relevant document is retrieved, then this value is averaged over the information needs. This metric is mathematically expressed in Eq.6:

$$\mathrm{MAP}(Q) = \frac{1}{|Q|}\sum_{j=1}^{|Q|}\frac{1}{m_j}\sum_{k=1}^{m_j}\mathrm{Precision}(R_{jk}) \tag{6}$$

where: m_j is the number of relevant documents at position j in the ranking and R_{jk} are the retrieved documents from the top of the ranking until the document k is reached.

3.3.3 Baselines

As reference, we have considered one naive baseline and three state-of-the-art results:

1. **Followers:** The ranking of authors according to their descending number of followers is the baseline of the RepLab 2014 competition. The number of followers is a basic indication of the author's authority potential.

2. **AleAhmad et al.** (AleAhmad et al., 2014). The main idea in this study is that influencers or opinion makers talk more about hot topics. This method extracts hot topics from each domain and a time-sensitive voting algorithm is used to rank each author on their respective topic.

3. **Rodríguez-Vidal et al.** (Rodríguez-Vidal et al., 2019). This approach got the best results in the RepLab 2014 competition. Here the authors applied language models of authorities and domain (automotive and banking) knowledge for identify and characterize influencers.

4 Results and discussion

Table 1 summarizes the results of all experiments for each signal explained in Section 3.1.1. Note that, for those experiments that only use a single signal, their results are presented for the *DSR* strategy only, since *L2R* behaves like *DSR* with one signal.

Regarding the results shown in Table 1, we can extract the following conclusions:

- The language used by the followers allows to characterize a profile much better than using only the number of followers provided by the Social Network (0.75 vs 0.38, an improvement of 97%).

- Regarding the results extracted from each signal individually, the best result

	DSR	L2R
#_Foll_Auth	0.39	-
Mod_Foll_Auth	0.44	-
Avg_Mod_Foll_Auth	0.47	-
Prop_Foll_Auth	0.44	-
Avg_Prob_Auth	0.41	-
Sum_Foll_Auth	0.42	-
#_Foll_Dom	0.35	-
Mod_Foll_Dom	0.41	-
Avg_Mod_Foll_Dom	0.42	-
Prop_Foll_Dom	0.44	-
Avg_Prob_Dom	0.42	-
Sum_Foll_Dom	0.45	-
All Combined	0.58	**0.61**
Followers	0.38	

Table 1: Signals from followers in isolation

in our experiments comes from the average authority of followers. This means that the number of followers is not as important as their average quality. This is relevant because the diffusion of a message through followers will be faster, and therefore will have a greater impact, if other influencers validate and spread that message.

- Nevertheless, combine all authority and domain signals, is the best way to use the followers' information.

Table 2 summarizes the results of combining the information extracted from the main profiles themselves with the information extracted from her followers, and compares these results against the information extracted exclusively from the followers and against other state-of-the-art systems.

	DSR	L2R
Main_Text + Followers_Text	0.74	**0.75**
Followers_Text	0.58	0.61
Rodríguez-Vidal et al.	0.68	0.74
AleAhmad et al.	0.57	
Followers	0.38	

Table 2: Adding followers' signals to the main profile

From the results shown in Table 2, we may conclude:

- The combination of the information provided by the main profile with that of her followers is a better option to locate influencers than using the information of the followers in isolation. We obtain an improvement of 18.66% for the L2R technique.
- The difference existing between the ranking techniques is irrelevant (1%), so the unsupervised approach (DSR) provides similar results than the supervised one (L2R).
- The addition of followers' signals improves overall results by a short margin (1.35%), which may indicate that the information of both actors is redundant.
- The short margin of improvement obtained and the effort taken for collecting and processing the information of the followers, lead us to think in the need to value the cost/benefit of using this approach according to the characteristics of the Social Network of study.

5 *Conclusions*

Our main goal in this study was to investigate the role of the followers in the task of finding Twitter influencers for a given domain. To do so, we model their language and extract different signals which help us to characterize authorities and domain experts and we compare our results against the state-of-the-art. The main conclusions of our experiments are:

- The followers of a user may provide useful information for characterizing her. The profiles followed by other influencers are more likely to be influencers. This indicates that influencers tend to be connected with other of their kind.
- This discovery leads us to an interesting discussion. Since influencers tend to be connected to each other, an idea written by one of them is accepted and validated (e.g. using a retweet) by other influencers, and this may cause more severe reputational crises since: (i) the diffusion of a message by other influencers adds a new audience to that idea; and (ii) if the opinions of one influencer are reliable for the users of a given community, the validation by another influencer(s) gives the message a greater veracity in the eyes of that community.
- The combination of the information provided by the main profiles with the information of her followers produces the

best results, but only by a short margin, which may indicate that these information is redundant.

Acknowledgments

This research was supported by the Spanish Ministry of Science and Innovation (Vemodalen Project, TIN2015-71785-R).

Bibliografía

AleAhmad, A., P. Karisani, M. Rahgozar, y F. Oroumchian. 2014. University of tehran at replab 2014. En *Conference and Labs of the Evaluation Forum (Working Notes)*, páginas 1528–1536.

Amigó, E., J. Carrillo-de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, y D. Spina. 2014. Overview of replab 2014: author profiling and reputation dimensions for online reputation management. En *International Conference of the Cross-Language Evaluation Forum for European Languages*, páginas 307–322. Springer.

Aral, S. y D. Walker. 2012. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341.

CLEF. 2014. Replab 2014 campaign @online. http://nlp.uned.es/replab2014/.

Cossu, J.-V., K. Janod, E. Ferreira, J. Gaillard, y M. El-Bèze. 2014. Lia@ replab 2014: 10 methods for 3 tasks. En *4th International Conference of the Conference and Labs of the Evaluation Forum initiative*. Citeseer.

Cossu, J.-V., V. Labatut, y N. Dugué. 2016. A review of features for the discrimination of twitter users: application to the prediction of offline influence. *Social Network Analysis and Mining*, 6(1):1–23.

Dang, V. 2012. The lemur project-wiki-ranklib. *Lemur Project,[Online]. Available:* `http://sourceforge.net/p/lemur/wiki/RankLib`.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, páginas 1189–1232.

Jefferson-Henrique. 2016. Getoldtweets-java @online. https://github.com/Jefferson-Henrique/GetOldTweets-java, Abril.

Liu, T.-Y. 2009. Learning to rank for information retrieval. *Found. Trends Inf. Retr.*, 3(3):225–331, Marzo.

Lomena, J. J. M. y F. L. Ostenero. 2014. Uned at clef replab 2014: Author profiling task. En *Conference and Labs of the Evaluation Forum (Working Notes)*, páginas 1537–1546.

Madden, M. y A. Smith. 2010. Reputation management and social media. *pewresearch.org,[Online]. Available:* `pewresearch.org/internet/2010/05/26/reputation-management-and-social-media/`.

Manning, C. D., P. Raghavan, y H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Nebot, V., F. Rangel, R. Berlanga, y P. Rosso. 2018. Identifying and classifying influencers in twitter only with textual information. En *International Conference on Applications of Natural Language to Information Systems*, páginas 28–39. Springer.

Rodríguez-Vidal, J., J. Gonzalo, L. Plaza, y H. A. Sánchez. 2019. Automatic detection of influencers in social networks: Authority versus domain signals. *Journal of the Association for Information Science and Technology*.

Saari, D. G. 1999. Explaining all three-alternative voting outcomes. *Journal of Economic Theory*, 87:313–355.

Vilares, D., M. Hermo, M. A. Alonso, C. Gómez-Rodríguez, y J. Vilares. 2014. Lys at clef replab 2014: Creating the state of the art in author influence ranking and reputation classification on twitter. En *Conference and Labs of the Evaluation Forum (Working Notes)*, páginas 1468–1478.

Villatoro-Tello, E., G. Ramírez-de-la Rosa, C. Sánchez-Sánchez, H. Jiménez-Salazar, W. A. Luna-Ramírez, y C. Rodríguez-Lucatero. 2014. Uamclyr at replab 2014: Author profiling task. En *Conference and Labs of the Evaluation Forum (Working Notes)*, páginas 1547–1558.