Universidad Pablo de Olavide (España)
Revista de Métodos Cuantitativos para la Economía y la Empresa
número 38, 2024
ISSN: 1886-516X
DOI: 10.46661/revmetodoscuanteconempresa.8135
Sección: Artículos
Recibido: 15-05-2023
Aceptado: 11-04-2024
Publicado: 03-12-2024
Páginas: 1-13

# Detección de factores relevantes en redes sociales incorporando información de expertos

## *Incorporating Expert Judgment for Detecting Relevant Factors in Social Networks Undetected by Ordinary Methods*

José María Pérez-Sánchez
Universidad de Las Palmas de Gran Canaria (España)
https://orcid.org/0000-0002-7491-4345
josemaria.perez@ulpgc.es

María Acosta García
Universidad de Granada (España)
ag.maria49@gmail.com

Emilio Gómez Déniz
Universidad de Las Palmas de Gran Canaria (España)
https://orcid.org/0000-0002-5072-7908
emilio.gomez-deniz@ulpgc.es

**RESUMEN**

La mayoría de las empresas utilizan las redes sociales como canales de comunicación debido a los beneficios empresariales que pueden proporcionar. Este artículo se centra en el impacto de las redes sociales en una fundación española dedicada a la innovación y la difusión del conocimiento, y en cómo afectan a sus principales eventos y actividades. Examinamos los factores subyacentes a un retuit en Twitter o una compartición en Facebook con el fin de analizar la difusión de los eventos principales de esta fundación. Se realizaron comparaciones con tres modelos estadísticos (regresión estándar y regresión bayesiana con distribución a priori informativa y no informativa). Concluimos que la ventaja que ofrece la metodología bayesiana sobre la clásica se demuestra mediante la incorporación de información colateral,



***Cómo citar:*** Pérez Sánchez, J. M., Acosta García, M., & Gómez-Déniz, E. (2024). Detección de factores relevantes en redes sociales incorporando información de expertos. *Revista De Métodos Cuantitativos Para La Economía Y La Empresa*, 38. https://doi.org/10.46661/rev.metodoscuant.econ.empresa.8135

generalmente proporcionada por expertos, lo que permite refinar el modelo y obtener conclusiones que, de otro modo, no podrían ser identificadas. Esta conclusión puede tener implicaciones significativas para las empresas que utilizan redes sociales.

**PALABRAS CLAVE**

Inferencia bayesiana; simulación MCMC; distribución a priori informativa; redes sociales.

**ABSTRACT**

Information and communications technology (ICT) has potential to complement information sharing bureaus (ISB) Most companies use social networks as communication channels because they can provide significant business benefits. This paper focuses on the impact of social networks in a Spanish foundation for innovation and knowledge dissemination, and how they affect its main events and activities. We examine the factors underlying a re-tweet on Twitter or a share on Facebook in order to analyze reporting of this foundation's principal events. Comparisons with three statistical models were performed (standard regression and Bayesian regression with non-informative and informative priors). We conclude that the advantage offered by Bayesian over classic methodology is demonstrated by incorporation of collateral information, usually provided by experts, which can refine the model and obtain conclusions that cannot be identified otherwise. This conclusion may have significant implications for companies that make use of social networks.

**KEYWORDS**

Bayesian inference; MCMC simulation methods; informative prior distributions; social networks.

JEL classification: C01, C11.

MSC2010: 62–08, 62P25.

## 1. INTRODUCTION

Although, teenagers today are the most prolific users of social networking sites (SNS), adults and private companies, as well as public, health, educational, etc. organizations, are also aware of their importance to highlight their roles and take advantage of the many possibilities they offer. Some of the main social networks are Facebook, YouTube, Whatsapp, Instagram, Google+ and Spotify, among others.

Companies increasingly use social networks because they consider them valuable communication channels. In fact, according to statistical data, in 2021 around 70% of Spanish companies use social networks to promote their brands, optimize communication and encourage consumers to know them better (Statista, 2022).

A couple of examples of the enormous power of social media to promote a brand and retain customers include the re-tweets of "Securitas Direct", a leading alarm systems company, on Twitter, which encourage customers to access content. Another case is that of "Remica", a company dedicated to energy efficiency, which through its "Canal de Remica" (Remica channel) advertises its services and sector-related news through videos uploaded to YouTube.

A further example relates to a Spanish political party, "Unidas Podemos", predominantly constituted by politically-motivated young people, which has used social networks to make themselves known and gain considerable electoral success. Following this, Barack Obama's team created profiles in the best-known social networks, and became in some ways pioneers in using these tools in his campaign for the US presidency. In this context, the Washington Post named him the King of Networks. On the other hand, everyone is familiar with the fondness that the ex-US President Donald Trump has for social media, who converted his stream of tweets into a nexus for debates both online and in traditional media. Other political parties have had no choice but to follow, so as not to be left behind.

This paper analyzes the importance of SNSs in a Spanish foundation whose main goal is the diffusion of knowledge and innovation. Specifically, we are interested in SNSs impact on its main events and activities, as measured by its number of followers on Twitter and Facebook. This is an approach that can easily be transferred to other scenarios, such as politics or financially successful companies.

There has been growing academic interest in SNS use, and, as such, the methodology for studying its impact on society is very diverse. Wasserman and Faust (1994) and Carrington et al. (2005), for example, presented excellent overviews of these topics. A brief summary of studies might include: Hsu et al. (2007), who conducted an analysis of latent and multidimensional spatial models for predicting, classifying, and annotating friends' relationships in networks, based upon network structures and user profile data. García de Torres et al. (2011), on the other hand, applied an exploratory study to examine the use of social media by combining the analysis of Twitter and Facebook profiles and semi-structured interviews. Túñez and Sixto (2011) investigated the types of information posted on Facebook pages of a sample of politicians. Gúzman et al. (2012) analyzed the top 20 Latin American universities that are active on Twitter by identifying their lists, fans and tweets, and re-tweets. Harrigan et al. (2012) employed a conditional logistic regression model in analyzing the re-tweeting behaviour of a community of Twitter users. Another interesting study was conducted by the University of Santiago de Compostela to assess the use of Facebook in university teaching (Túñez and Sixto, 2012), which proved acceptance of its use as a learning tool in the university context. Golbeck and Hansen (2014) presented an alternative method for computing political preferences of users on Twitter based on certain behaviour. Cluster analysis, kmeans and multivariate analysis were other mechanisms used by Parra et al. (2014) for analyzing the usefulness of social networks in the socialdemographic and cultural characterization of risk groups. More recently, Eriksson and Olsson Gardell (2016) analyzed the role of social networks among citizens and crisis communication professionals.

Few studies about SNSs, however, have been published by estimating from a Bayesian perspective. Butts (2003) obtained posterior simulation for models with the presence of measurement error and missing data using Markov Chain Monte Carlo (MCMC) methods. Koskinen and Snijders (2007) proposed an alternative Bayesian estimation procedure to the method of moment estimation for longitudinal social network data. Others works have applied Bayesian inference in the study of exponential random graph models (Caimo and Friel, 2011; 2013; Koskinen et al., 2013 and Slaughter and Koehly, 2016; among others).

The advantage offered by Bayesian over classic methodology is shown by the incorporation of collateral information, usually provided by experts, that can refine the model and obtain conclusions that cannot otherwise be seen. In this work, this is evidenced by the fact that certain factors (that classical models are unable to capture as influential in the number of shares/re-tweets) are identified under the Bayesian methodology as significant. This conclusion might be important to companies that make use of social networks. For example, a company may obtain the opinion of potential customers by analyzing the shares/re-tweets in relation to new products.

The remainder of this paper is structured as follows. Section 2 describes the foundation database, which have been assessed in this work. The models proposed for fitting the influence of events and other factors are presented in Section 3; while Section 4 discusses the results, and investigates specific diagnostic measures to decide on the best model. Finally, Section 5 presents the concluding remarks and discussion.

## 2. DATABASE

Here we analyze the importance of SNSs in an organization that diffuses knowledge and innovation. As the focus is on how SNSs impact events and activities, as said, number of followers on Twitter and Facebook will be used as a measure. We begin this section by describing the data source. The data are drawn from a Spanish private non-profit foundation for the dissemination of innovation, backed by the regional government of Andalusia, that works with more than 22 institutions, which are public universities, research and dissemination centres. The aim of the foundation is to foster cultural knowledge in the population through organizing, coordinating and driving initiatives for the dissemination of innovation, science and technology. Facebook data were collected in a time series study of 54 weeks from April 2012 to April 2013 and 105 weeks from December 2011 to December 2013 for Twitter.

### 2.1. Facebook data

The dependent variable for the Facebook study is the number of “shares”, defined by the publication of any news from the foundation by followers on their wall. In total, six explanatory variables are considered: “followers” represents the difference between the number of people or institutions that follow the foundation’s Facebook page from one week to the previous; “friends” measures the difference between the number of people or institutions that follow the foundations Facebook page from one week to the previous; and finally, “visits” represents the number of visits to the foundations Facebook page. We have created three dichotomous explanatory variables which represent the most important events organized by the foundation in this period. These variables equal 1 for the week of the event and 0 for the remaining weeks. “science12” represents the Science and Innovation week (second week of November 2012), an event in which during two weeks many firms related to research and knowledge organize a wide range of outreach science events. “night12” is Researchers Night (fourth week of September 2012), another event in which, for a night, several researchers organize activities for the dissemination of innovation, science and technology. And finally, we consider the event Womens Coffee with Science (second week of March 2013), denoted by “wcoffee13”, an activity organized by the foundation in which female researchers discuss current research topics. Table 1 shows a descriptive summary of the variables considered for the Facebook model including means, standard deviances, minima and maxima.

**Table 1. Variables and descriptive summary. Facebook model.**

| Variables | Mean | s.d. | Min. | Max. |
|---|---|---|---|---|
| shares | 180.35 | 101.24 | 23 | 712 |
| followers | 24.91 | 16.46 | 1 | 88 |
| friends | 9697.15 | 8589.09 | -475 | 42369 |
| visits | 4205.94 | 1831.45 | 818 | 12949 |
| science12 | 0.44 | 0.50 | 0 | 1 |
| night12 | 0.57 | 0.49 | 0 | 1 |
| wcoffee13 | 0.13 | 0.34 | 0 | 1 |

### 2.2. Twitter data

The dependent variable for the Twitter study is the number of “re-tweets”; defined by the publishing of any tweets by the foundation’s followers. In total, nine explanatory variables are considered: “tweets” relates to the number of twitter posts from the foundation; “followed” represents the difference between people or institutions followed by the foundation from one week to the previous; “followers” shows the difference between people or institutions that follow the foundation from one week to the previous; and finally, "mentions" measures the number of mentions of the foundation user-name by another user. For this study, we have used five dichotomous explanatory variables, which represent the most important events organized by the foundation in this period; and again they equal one from the week of the event. These events are the same as the Facebook study but, in this case, we have two variables for the Science and Innovation week (second week of November 2012 and 2013) and another two for Researchers Night (fourth weeks of September 2012 and 2013), because the sample contains two years. Table 2 shows the descriptive summary for Twitter variables.

**Table 2. Variables and descriptive summary. Twitter model.**

| Variables | Mean | s.d. | Min. | Max. |
|---|---|---|---|---|
| re-tweets | 35.98 | 26.54 | 4 | 163 |
| tweets | 38.43 | 27.11 | 2 | 162 |
| followed | 1.84 | 2.82 | −2 | 15 |
| followers | 33 | 15.89 | −11 | 89 |
| mentions | 15.17 | 12.90 | 1 | 82 |
| science12 | 0.54 | 0.50 | 0 | 1 |
| science13 | 0.06 | 0.23 | 0 | 1 |
| night12 | 0.61 | 0.49 | 0 | 1 |
| night13 | 0.12 | 0.33 | 0 | 1 |
| wcoffee13 | 0.40 | 0.49 | 0 | 1 |

## 3. METHODOLOGY

We propose two alternative multiple linear regression models to detect relevant factors on social networks. Frequentist and Bayesian estimation approaches have been used in order to compare the results. From the Bayesian perspective, we have considered both non-informative and informative prior distributions. We will have the opportunity to verify that the latter is the model that best fits the data, incorporating factors as significant that do not appear as influential in the previous models.

Let $\mathrm{y} = (y_1, y_2, \ldots, y_n)'$ denote an $n \times 1$ vector of a dependent variable. We assume that $y_i\,(x_i, \beta)$ depends on a vector $k$ of unknown regression coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ and then consider the function $h(x_i) = \sum^k x_{ij}\beta_j$, where $h(\cdot)$ is the score function. Here, $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})'$ denotes the $k \times 1$ vector of covariates for the $i$ week. A regression model deals with the problem of estimating the variable $y_i$ and is given by

$$y_i = F(x_i' \boldsymbol{\beta}) + \varepsilon_i, \qquad (1)$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients, which represents the effect of each variable in the model and $F(\cdot)$ is the link function and $\boldsymbol{\varepsilon}_i = (\varepsilon_1, \ldots, \varepsilon_n)'$ is the error term assumed to follow a multivariate normal distribution with mean $\mathbf{0}_n$ and covariance matrix $\sigma\ 2\boldsymbol{I}_n$.

### 3.1. Frequentist estimation of the regression models

For conventional linear regression models, the link function is equal to $F(\mu_i) = \sum^k \beta_j x_{i\ j}$, for $j = 1, \ldots, k$ and $\mu_i = E(y_i|x_i)$. The conditional mean for the $i$-th response can be written as $\mu_i = E(y_i|x_i) = \mu(x_i, \beta)$, $i = 1, 2, \ldots, n$. Therefore,

$$\mu_i = E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i,\ i = 1, \ldots, n. \quad (2)$$

The regression coefficients, βi, are usually estimated by ordinary least squares (OLS). We have estimated these coefficients by employing the OLS method under robust standard deviations for solving potential problems of autocorrelation and/or heteroscedasticity (Huber, 1967 and White, 1980, 1982). The econometric software package used here was STATA, although other packages such as Mathematica, Matlab and R are also suitable for this purpose.

### 3.2. Bayesian estimation of the regression models

Use of the Bayesian methodology in Social Sciences and other disciplines is frequent. This is reflected in the numerous research articles and monographs on the subject. These include Zellner (1996), Koop (2003) and Kaplan (2014), among others. In the social network scenario, the use of the Bayesian approach has proliferated in recent years due to the flexibility that it provides to model the essential networks and the nodes that interconnect them. In addition, the possibility of using more powerful computers to calculate and process large databases allows researchers to handle algorithms that were unthinkable in the past, such as MCMC, which has led to it being commonly used in the social network environment. Recent works on the subject include Caimo and Friel (2011; 2013) and Oliveira et al. (2017); among others.

In Bayesian statistics, uncertainty regarding unknown parameters is modelled by assigning a probability density function to the parameters of interest. Therefore, these are regarded as random variables. Next, the focus of interest in Bayesian inference is that the posterior distribution derives from the combination of the likelihood function with the prior density function that reflects previous knowledge on the distribution of the parameters of interest by using the Bayes rule. Thus, $\pi(\Theta|x) \propto l(x|\Theta)\pi(\Theta)$, where $\Theta$ is the parameter (vector of parameters) of interest, $l(x|\Theta)$ the likelihood and $\pi(\Theta|x)$ is the posterior distribution and tells us how the parameter is distributed after the data, x, have been observed. See, for example Zellner (1996) and Kaplan (2014).

From the Bayesian point of view, the model is expressed according to the following schedule,

$$y_i \sim N(\mu_i,\ \tau), \quad (3)$$

where μi the average number of shares/re-tweets and $\tau$ is the precision, i.e., the inverse of the variance, $\sigma$ 2. Under Bayesian assumptions, μi = F(xi' β) and due to the form of the likelihood function in the OLS model, it would suggest that the natural conjugate priors for the hyperparameters $\beta\,|\tau$ and $\tau$ are considered to be random variables with normal and gamma distribution, respectively (see for instance, Koop (2003), Chapter 3, p. 36). That is,

$$\boldsymbol{\beta}\,|\tau \sim N(\mu_{\boldsymbol{\beta}},\ \tau_{\boldsymbol{\beta}}), \quad (4)$$

$$\tau \sim G(a,\ b). \quad (5)$$

Therefore, $\pi(\boldsymbol{\beta},\ \tau) = \pi(\boldsymbol{\beta}\ |\tau)\pi(\tau)$ follows a normal-gamma distribution. This joint distribution together with the likelihood is needed to get the posterior. Now, we compute the posterior distribution of $\mu_i$, assuming the hierarchical structure given by (3), (4) and (5). Some algebra confirms that the posterior is not recognize as a standard distribution. Thus, we get a complicated posterior which does not lead to convenient expressions for the marginals of $\boldsymbol{\beta}$ and $\tau$ and computational methods are required. The most popular numerical method, the Gibbs Sampler, uses the conditional posteriors.

We can sample $\boldsymbol{\beta}$ from the posterior distribution by using a WinBUGS package (Windows Bayesian inference Using Gibbs Sampling, developed jointly by MRC Biostatistics Unit (University of Cambridge, Cambridge, UK) and the Imperial College School of Medicine at St. Marys, London, UK); see Lunn et al. (2000); based on Gibbs sampling applying MCMC methods (see Carlin and Polson, 1992, and Gilks et al., 1995, for further details).

### 3.2.1. Non-informative prior distributions

An important problem in Bayesian analysis is the elicitation process, i.e., how to define the prior distributions which include the prior knowledge. If there is a total lack of prior belief in the Bayesian estimator, the estimator becomes a function of the likelihood. In this case, we assume centred and non-informative prior distributions to facilitate comparison with frequentist methods of estimation by considering that

$$\beta_j \sim N(0,\ 10^{-6}),\ \ \forall j = 0,\ \dots,\ k,$$
$$\tau \sim G(10^{-3},\ 10^{-3}). \tag{6}$$

In this context, for the $\boldsymbol{\beta}$ coefficients, we are assuming that their precisions (variances) are very small (high) allowing a wide interval for the estimations of the parameters. In the other hand, by assuming a gamma distribution with $a = 10^{-3}$ and $b = 10^{-3}$, we are supposing a non-informative distribution for $\tau$. So, no prior knowledge exists about the coefficients of the model and the precision.

### 3.2.2. Informative prior distributions

Bayesian researchers traditionally consider the prior probability distribution over the parameter or parameters of interest, before the data is collected, and which represents the state of the nature of the problem that concerns them. However, there is another way of acting, also within the Bayesian statistic, that allows information considered useful, although not exhaustive, to be transmitted by the researcher, or by an expert to whom an opinion is requested, and that allows us to specify in a better way a prior distribution. Sometimes an informative prior distribution is said to dominate the likelihood. This will be the line of action in this section.

Thus, the methodology developed in this section allows us firstly to incorporate the knowledge of an expert to elicit the hyper-parameters of the developed model, and assign the probability distribution. This expert opinion prior to developing Bayes theorem allows a more flexible and coherent model to be obtained, which will lead, as will be seen in the example provided below, to the significance of certain factors that are not detectable by the frequentist and non-informative models.

Thus, we demonstrate the performance of the Bayesian regression model in the setting considered here. In comparison with the previous section, we asked an expert from the foundation about the expected impact of the events considered in this study in the number of shares/re-tweets. The remaining prior distributions are considered non-informative, as in the previous analysis. Next, expected increases are obtained and corresponding normal distributions are assumed, as can be seen in Table 3. In this Table, the experts' expected increases are approximated to a Normal distribution with the mean equal to the expected value. Furthermore, high precisions (small variances) are considered in order to assume informative prior information. We have assumed higher precisions for distributions that represent the number of re-tweets than for the number of shares, in order to take into account the lower range of variability of this dependent variable (see Tables 1 and 2).

N. 38, 2024 - ISSN: 1886-516X - DOI: 10.46661/revmetodoscuanteconempresa.8135 - [Págs. 1-13]



**Table 3. Elicitation process for Facebook and Twitter.**

| | Shares | | re-tweets | |
|---|---|---|---|---|
| **Variable** | **Expected** | **Distribution** | **Expected** | **Distribution** |
| science12 | 10 | $N\,(10, 10^{-2})$ | 10 | $N\,(10, 10^{-1})$ |
| science13 | – | – | 10 | $N\,(10, 10^{-1})$ |
| night12 | 40 | $N\,(40, 10^{-2})$ | 20 | $N\,(20, 10^{-1})$ |
| night13 | – | – | 20 | $N\,(20, 10^{-1})$ |
| wcoffee13 | 10 | $N\,(10, 10^{-2})$ | 10 | $N\,(10, 10^{-1})$ |

# 4. RESULTS

The statistical methods consisted in two steps:

i) Estimation of the regression models and analysis of the relevance variables.

ii) Validation of goodness of fit using Akaike information criterion (AIC) for frequentist estimation and deviance information criterion (DIC) for Bayesian estimation. Note that AIC $= 2(k - l_{max})$, where $k$ is the number of model parameters and $l_{max}$ is the maximum value of the log–likelihood function, and $DIC = -2\log(l(y|x, \beta))$. Both statistics measure the relative quality of statistical models for a given set of data. The idea is that models with smaller AIC and DIC should be preferred to models with larger AIC and DIC. See Akaike (1974) and Spiegelhalter et al. (2002) for details.

The final sample sizes were 54 and 105 weeks for the Facebook and Twitter studies, respectively. For each model, we have estimated two versions: full and reduced. The former includes all covariates and the latter only considers those covariates that turned out to be relevant (p-value smaller than 10%) in the full model.

## 4.1. Facebook regressions models

### 4.1.1. Frequentist estimation of the regression model

In this model, the dependent variable is the number of shares. Table 4 includes the estimations of the parameters and the robust standard errors for the full and reduced models. A summary with the sample size and AIC is also included. The first model is the full model, including all covariates. The AIC is 584.624 and the visits is found to be positive and significant at 1% in explaining the dependent variable. Furthermore, the Researchers Night of 2012 led to a significant increase in the number of shares (at 10% significance). So, in estimating the reduced model, we only consider these two significant variables (and the intercept). The AIC, in this case, is 577.533, indicating a best fit, and two considered variables remain significant with the same level.

**Table 4. Frequentist estimation of Facebook model.**

| Model | Full | | Reduced | |
|---|---|---|---|---|
| Variable | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE |
| intercept | −34.467 | 32.979 | −38.513 | 25.878 |
| followers | −0.325 | 0.614 | | |
| Friends | $-2 \cdot 10^{-5}$ | $6 \cdot 10^{-4}$ | | |
| visits | 0.049*** | 0.008 | 0.048*** | 0.007 |
| science12 | −10.832 | 21.124 | | |
| night12 | 36.919* | 18.443 | 27.931* | 14.354 |
| wcoffee13 | 4.501 | 21.608 | | |
| *n* | 54 | | 54 | |
| AIC | 584.624 | | 577.533 | |

**Table 5. Bayesian estimation of Facebook model.**

| Model | Non-informative | | | | Informative | | | |
|---|---|---|---|---|---|---|---|---|
| | Full | | Reduced | | Full | | Reduced | |
| Variable | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE |
| intercept | −34.47 | 23.33 | −38.5*** | 19.38 | −52.18*** | 19.3 | −51.95*** | 19.12 |
| followers | −0.324 | 0.442 | | | 0.236*** | 0.202 | 0.236*** | 0.203 |
| friends | $-2 \cdot 10^{-5}$ | $8 \cdot 10^{-4}$ | | | $6 \cdot 10^{-4}$*** | $5 \cdot 10^{-4}$ | $6 \cdot 10^{-4}$*** | $5 \cdot 10^{-4}$ |
| visits | 0.049*** | 0.004 | 0.048*** | 0.003 | 0.047*** | 0.004 | 0.047*** | 0.004 |
| science12 | −10.79 | 25.02 | | | 2.81 | 8.564 | | |
| night12 | 36.85* | 23.83 | 27.94*** | 13.9 | 34.48*** | 8.485 | 35.73*** | 8.159 |
| wcoffee13 | 4.538 | 23.47 | | | 7.605 | 9.065 | | |
| *n* | 54 | | | | 54 | | | |
| DIC | 587.409 | | 579.756 | | 583.277 | | 578.059 | |

### 4.1.2. Bayesian estimation of the regression model

Table 5 shows the results from Bayesian estimations. For the non-informative model, the relevant variable at 1% of relevance is the number of visits, as in the frequentist model. Again, the Researchers Night of 2012 was an important event, which increased the number of shares significantly. Remaining explicative variables are not relevant in order to explain the number of shares. On the other hand, in the informative model the results change increasing the number of relevant variables: the number of followers, friends and visits are all positive and relevant factors at 1% in explaining the number of shares. Finally, the Science and Innovation week in 2012 and the Womens Coffee with Science in 2013 did not significantly influence the number of shares. Reduced models concluded in the same way for both non-informative and informative Bayesian estimations. The lower DIC is obtained by the informative reduced model with 578.059.

## 4.2. Twitter regression models

### 4.2.1. Frequentist estimation of the regression model

Estimations for the number of re-tweets from a frequentist point of view are shown in Table 6. We observe that the significant variables are the number of tweets and mentions, and the 2013 Researchers Night event. The number of tweets and the Researchers Night of 2013 made the number of re-tweets increase (with a 1% level of significance). On the other hand, the number of mentions decreases the number of re-tweets (at 1%), noting that mentions are a substitute option for re-tweets. The reduced model obtains the lower AIC, i.e., 874.559.

**Table 6. Frequentist estimation of Twitter model.**

| Model | Full | | Reduced | |
|---|---|---|---|---|
| Variable | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE |
| intercept | 5.850 | 4.307 | 10.290*** | 3.357 |
| tweets | 0.734*** | 0.094 | 0.776*** | 0.097 |
| mentions | -0.497*** | 0.245 | -0.500*** | 0.218 |
| followers | 0.197 | 0.118 | | |
| followed | -0.558 | 0.655 | | |
| science12 | -12.215 | 8.770 | | |
| science13 | 14.432 | 7.426 | | |
| night12 | 10.432 | 7.426 | | |
| night13 | 22.285*** | 7.633 | 27.804*** | 6.618 |
| wcoffee13 | 1.415 | 4.741 | | |
| *n* | 105 | | 105 | |
| AIC | 877.745 | | 874.559 | |

### 4.2.2. Bayesian estimation of the regression model

Table 7 shows the results from Bayesian estimations. With respect to the non-informative Bayesian estimations, the number of tweets, the Researchers Night in 2013 (both at 1% of relevance) and the number of followers (at 10%) are positive and relevant factors that explain the number of re-tweets. On the other hand, the number of mentions (at 1%) and the Science Week in 2012 (at 10%) are negative and relevant variables. The reduced model only detects relevance on the number of tweets, mentions and the Researchers Night 2013 event with the same signs, as expected. The informative Bayesian estimations find new relevant factors apart from those already detected by the non-informative estimations. The number of people the company follows (followed), Science Week 2013 and the Researchers Night 2012 events are relevant when the experts knowledge is taken into account. Furthermore, the number of followers is again relevant (now at 1%). The informative reduced model highlights the robustness of the previous and new results. Again, the lower DIC is obtained by the reduced model.

N. 38, 2024 - ISSN: 1886-516X - DOI: 10.46661/revmetodoscuanteconempresa.8135 - [Págs. 1-13]



**Table 7. Bayesian estimation of Twitter model.**

| Model | Non-informative | | | | Informative | | | |
|---|---|---|---|---|---|---|---|---|
| | Full | | Reduced | | Full | | Reduced | |
| Variable | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE | $\hat{\beta}$ | SE |
| intercept | 5.849 | 4.481 | 7.49** | 4.275 | 3.31 | 4.152 | 2.789 | 4.094 |
| tweets | 0.734*** | 0.067 | 0.751*** | 0.066 | 0.729*** | 0.065 | 0.743*** | 0.064 |
| mentions | −0.497*** | 0.143 | −0.509*** | 0.145 | −0.481*** | 0.139 | −0.471*** | 0.141 |
| followers | 0.197* | 0.109 | 0.122 | 0.104 | 0.153** | 0.087 | 0.136*** | 0.083 |
| followed | −0.558 | 0.665 | | | 0.389** | 0.322 | 0.406*** | 0.334 |
| science12 | −12.21* | 7.233 | −0.603 | 3.416 | −7.825 | 5.394 | | |
| science13 | 14.08 | 8.622 | | | 12.09* | 6.294 | 11.78** | 6.317 |
| nigth12 | 10.43 | 6.616 | | | 10.64** | 4.965 | 5.387** | 3.229 |
| nigth13 | 23.29*** | 7.067 | 29.37*** | 5.778 | 21.92*** | 5.463 | 20.96*** | 5.259 |
| wcoffee13 | 1.417 | 4.953 | | | 1.341 | 4.249 | | |
| *n* | 105 | | | | 105 | | | |
| DIC | 880.473 | | 879.368 | | 878.145 | | 871.154 | |

## 5. CONCLUDING REMARKS AND DISCUSSION

In this work we have analyzed three estimation methods of a regression model that sought to identify the factors that influence the number of shares/re-tweets on Facebook and Twitter, respectively. It is concluded that, at least for the database considered here, the Bayesian method, based on the use of prior information provided by experts, is more adequate than the classical and non-informative Bayesian regression models.

Under the Bayesian informative estimations, there are several important factors that a company has to take into consideration if this firm wants to increase its number of shares on Facebook. The number of people who follow the company on Facebook, the number of friends who those followers have and the number of visits to the company´s Facebook page are all positive and significant features. Additionally, Researchers Night 2012 was an important event that significantly increased the number of shares. Regarding Twitter, the number of tweets, the number of followers and the number of people who are followed by the company are positive important factors in determining the number of re-tweets. On the other hand, there is only one significative negative factor, i.e. the number of mentions. It seems logical to suggest that the more mentions there are, the less re-tweets there will be, because both can be substitute actions. Finally, three events were important in the number of re-tweets: the 2013 Science and Innovation week and the 2012 and 2013 Researchers Nights.

These Facebook and Twitter findings can be very important for companies that use social networks. Social media is a key communication channel for companies, allowing them to connect with their audience, improve their brand image, and increase their presence in the market effectively and profitably. In this sense, social media platforms are essential tools for any company nowadays, as they allow reaching a massive and diverse audience immediately and directly. Through digital platforms, companies can interact with their customers, receive real-time feedback, and adapt their products and services according to the needs and preferences of the target audience. Furthermore, social media is an excellent way to increase the visibility of the company, generate traffic to its website, and improve its positioning in search engines. On the other hand, social media is also an effective way to build and maintain a positive reputation for the company. Whether through the creation of relevant and high-quality content or the proper handling of image crises, social media allows companies to communicate transparently and authentically with their customers, which can strengthen trust in the brand and generate loyalty among consumers. For example, for those that are interested in the opinion of their potential clients (this opinion is measured by shares/re- tweets) on certain characteristics of the product they trade, if it is a financially successful company; or for any other relevant organization, such as political parties, who may be interested in the opinion of their potential voters or followers. We believe that these results provide useful information that can increase companies' chances of a share/re-tweet.

## FUNDING

JMPS and EGD were partially funded by grant PID2021-127989OB-I00 (Ministerio de Economía y Competitividad, Spain) and EGD was partially funded by grant TUR-RETOS2022-075 (Ministerio de Industria, Comercio y Turismo).

## REFERENCIAS

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Butts, C. (2003). Network inference, error, and informant (in)accuracy: a Bayesian approach. *Social Networks*, 25(2), 103–140.

Caimo, C. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1), 41–55.

Caimo, C. and Friel, N. (2013). Bayesian model selection for exponential random graph models. *Social Networks*, 35(1), 11–24.

Carlin, B.P. and Polson, N.G. (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. In: Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. eds. *Bayesian Statistics*, 4, 577–86. University Press, Oxford.

Carrington, P.J., Scott, J. andWasserman, S. (Eds.) (2005). *Models and Methods in Social Networks Analysis*. Cambridge University Press, Cambridge.

Eriksson, M. and Olsson Gardell, E. (2016). Facebook and Twitter in Crisis Communication: A Comparative Study of Crisis Communication Professionals and Citizens. *Journal of Contingencies and crisis Management,* 24(4), 198–208.

García-de-Torres, E., Yezerska, L., Rost, A., Calderón, M., Edo, C., Rojano, M., Said-Hung, E., Jerónimo, P., Arcila, C., Serrano-Tellería, A., Sánchez-Badillo, J. and Corredoira, J. (2011). Uso de Twitter y Facebook por los medios iberoamericanos. *El profesional de la información*, 20(6), 611–620.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1995). Introducing markov chain montecarlo. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J., eds. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Golbeck, J. and Hansen, D. (2014). A method for computing political preference among Twitter followers. *Social Networks*, 36, 177–184.

Guzmán, A.P., Moral, E. and González, F. (2012). Twitter use in the Latin American Universities. *RELATEC, Revista Latinoamericana de Tecnología Educativa*, 11(1), 27–39.

Harrigan, N., Achananuparp, P. and Lim, E. (2012). Influentials, novelty, and social contagion. The viral power of average friends, close communities, and old news. *Social Networks*, 34, 470–480.

Hsu, H.W., Lancaster, J., Paradesi, M. and Weninger, T. (2007). Structural Link Analysis from User Profiles and Friends Networks: A Feature Construction Approach. En *Proceedings of ICWSM-2007*.

Huber, P.J. (1967). The behavior of maximum likelihood estimate under nonstandard conditions. En *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221–233. Ed. Berkeley, University of California Press.

Kaplan, D. (2014). *Methodology in the social sciences. Bayesian statistics for the social sciences*. Guilford Press, New York.

Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons Inc., United Kingdom.

Koskinen, J.H. and Snijders, T.A.B. (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, 137, 3930–3938.

Koskinen, J.H., Robins, G.L., Wang, P. and Pattison, P.E. (2013). Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, 35(4), 514–527.

Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS: a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing*, 10, 325–337.

Oliveira, S.C., Cobre, J. and Ferreira, T.P. (2017). A Bayesian approach for the reliability of scientific co-authorship networks with emphasis on nodes. *Social Networks*, 48, 110–115.

Parra, P., Gordo, AJ. and D'Antonio, SA. (2014). Social research applied to social networks. A methodological innovation for the analysis of Facebook Likes. *Revista Latina de Comunicación Social*, 69, 195–212.

Slaughter, A.J. and Koehly, L.M. (2016). Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Social Networks*, 44, 334–345.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, Series B, 64(4), 583–639.

Statista (2022). Avaliable online: https://es.statista.com/estadisticas/698728/porcentaje-deempresas-que-hicieron-uso-de-las-redes-sociales-en-espana/

Túñez, M. and Sixto, J. (2011). Social networks, political and Commitment 2.0: Spanish deputies in Facebook. *Revista Latina de Comunicación Social*, 66, 1–25.

Túnez, M. and Sixto, J. (2012). The social networks like learning environment: analysis of facebook use in the university teaching. *Revista de Medios y Educación*, 41, 77–92.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

White, H.L., Jr. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817– 838.

White, H.L., Jr. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons Inc., United Kingdom.