Determinants of ResearchGate (RG) Score for the Top100 of Latin American Universities at Webometrics

Carolina Henao-Rodríguez[1](✉), Jenny-Paola Lis-Gutiérrez[2], Mercedes Gaitán-Angulo[2], Carmen Vásquez[3], Maritza Torres[4], and Amelec Viloria[5]

[1] Corporación Universitaria Minuto de Dios, Bogotá, Colombia
linda.henao@uniminuto.edu

[2] Fundación Universitaria Konrad Lorenz, Bogotá, Colombia
{jenny.lis,mercedes.gaitana}@konradlorenz.edu.co

[3] Universidad Nacional Experimental Politécnica "Antonio José de Sucre", Barquisimeto, Venezuela
cvasquez@unexpo.edu.ve

[4] Universidad Centroccidental "Lisandro Alvarado", Barquisimeto, Venezuela
mtorres@ucla.edu.ve

[5] Universidad de la Costa, Barranquilla, Colombia
aviloria7@cuc.edu.co

Abstract. This paper has the purpose of establishing the variables that explain the behavior of ResearchGate for the Top100 Latin American universities positioned in Webometrics database for January 2017. For this purpose, a search was carried out to get information about postgraduate courses and professors at the institutional websites and social networks, obtaining documents registered in Google Scholar. For the data analysis, the econometric technique of ordinary least squares was applied, a cross-sectional study for the year 2017 was conducted, and the individuals studied were the first 100 Latin American universities, obtaining a coefficient of determination of 73.82%. The results show that the most significant variables are the number of programs, the number of teacher's profiles registered in Google Scholar, the number of subscribers to the institutional YouTube channel, and the GDP per capita of the university origin country. Variables such as (i) number of undergraduate programs, (ii) number of scientific journals; (iii) number of documents found under the university domain; (iv) H-index of the 1st profile of researcher at the university; (vi) number of members of the institution; (v) SIR Scimago ranking of Higher Education Institutions; (vi) number of tweets published in the institutional account; (vii) number of followers in the Twitter institutional account; (vii) number of "likes" given to the institutional count, were not significant.

Keywords: ResearchGate · Universities · Google scholar · Academic assessment · Webometrics

Y. Tan and Y. Shi (Eds.): DMBD 2019, CCIS 1071, pp. 320–327, 2019.
https://doi.org/10.1007/978-981-32-9563-6_33

1 Introduction

When discussing the impact of the publications, it is important to review the issue of bibliometric indicators, such as the ResearchGate metrics (RG). ResearchGate is a high-impact academic network that was founded in 2008 by Ijad Madisch, Sören Hofmayer, and Horst Fickenscher. During the first half of 2018, ResearchGate has more than 15 million members [1], and contains important academic networking tools, with a wide catalog of bibliometric indicators, among which, ResearchGate Score stands out [2]. The ResearchGate Score is the flagship indicator calculated using an undisclosed algorithm to measure the scientific reputation [3].

This paper aims to establish the variables that explain the behavior of ResearchGate for the Top100 Latin American universities positioned in Webometrics database in 2017. An ordinary least square model was applied with ResearchGate Score as the dependent variable and with the following explanatory variables: (i) number of post-graduate programs, (ii) number of teacher's profiles registered in Google Scholar, (iii) number of subscribers to the institutional YouTube channel, and (iv) GDP in each country where the institution is located.

The topic of ResearchGate has been treated by different authors, like [4] who indicated that RG is a research-oriented academic social network that reflects the level of research activity in the universities. The study suggests that academic social networks act as indicators in the assessment of research activities and may be useful and credible for acquiring scholar resources, staying informed about research results, and promoting the academic influence. In the same way, [2] conducted a study in which the RG Score was analyzed and revealed the main advantages and disadvantages of this indicator, concluding that it does not measure the prestige of the researchers but their level of participation in the platform.

From another perspective, [5] conducted a research that allowed to assess if the data of use and publication of ResearchGate reflected the existing academic hierarchies. This study concludes that the classifications based on the ResearchGate statistics correlate moderately well with other classifications of academic institutions, suggesting that the use of ResearchGate broadly reflects the traditional distribution of academic capital. At the same time, [6] presents a method to capture the structure of a full scientific community (the community of Bibliometrics, Scientometrics, Informetrics, Webometrics, and Altmetrics) and the main agents that are part of it (scientists, documents, and sources) through the lens of Google Scholar Citations (GSC). The method was applied to a sample of 814 researchers in Bibliometrics with a public profile created in Google Scholar Citations, and later used in the other platforms, collecting the main indicators calculated by each of them. The results obtained from this study were: (i) ResearchGate indicators, as well as the readers of Mendeley, present a high correlation with all indicators of GSC; and (ii) there is a moderate correlation with the indicators in ResearcherID.

Regarding the use of the Webometrics data, there are several researches that have already used them. For example, [7, 8] compared different rankings like the Ranking of Shanghai, QS World University Ranking, SCimago Institutions Rankings SIR, and the

Web Ranking of Universities-Webometrics, finding that the indicators associated with the research and institutional capacity stand out as common criteria in the reviewed evaluation methodologies. This research firstly states that Brazil occupies the first positions in the four rankings; and secondly, that there is a greater number of Latin American universities in QS (40%) and Webometrics (31%), while in the other two rankings it does not exceed 8%. [9] carried out a cluster analysis of the Top100 Latin American universities positioned in the Webometrics database in January 2017. The research included information about postgraduate programs and social networks on the web sites of these institutions and teachers. The variable with the highest correlation with the ranking is the number of postgraduate programs.

In the same way, the study of [10], analyzed a group of manageable visibility factors corresponding to the universities present in the Top100 of the Webometrics database of Latin America published in January 2017 for the identification of profiles. For this purpose, data was collected about the academic offer and scientific journals published on each university website; figures on documents and profiles found in Google Scholar; activity on social networks; and the institutional score reported by ResearchGate as a scientific network. Clusters were formed by quartiles to characterize the visibility profiles of Latin American universities considering the variables studied. The higher offer of postgraduate programs and the presence in scientific networks and Google Scholar characterize the best positioned universities.

For its part, [8] and [10], used the Data Envelopment Analysis (DEA) for processing the academic data published on the website of each university, the content and profiles shown in Google Scholar (GS), the data published by the university in ResearchGate as scientific network, and finally, the data of social networks such as the Twitter and Facebook accounts of the corresponding institutions. The authors found that the postgraduate offer, visibility in GS, and the use of scientific and social networks contribute favorably to the web positioning of Latin American universities.

From another approach, [11], also discussed the importance of webmetrics techniques for measuring the visibility, specifically in the case of university libraries in Sri Lanka. Similarly, [12] uses two quantitative techniques in Multi-criteria Decision Analysis (MCDA) which are the Entropy method and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), applied to the Webometrics ranking for universities in the world. These models help evaluators to apply a strategic vision for future developments by the use of the multi-criteria decision analysis method. The author concludes that Webometrics classification systems are perceived differently by different stakeholders and, therefore, can be approached in different ways.

2 Method

For the development of the document, an econometric exercise was performed to analyze the determinants of the ResearchGate (RG) score, since it is an indicator that considers the popularity and commitment to the RG community [13]. Therefore, it measures the number of publications, followers, and interactions within this scientific social researcher's network.

The econometric technique used was the ordinary least squares. A cross-section was made for the year 2017, where the individuals studied were the first 100 Latin American universities of the Webometrics Ranking for that year.

2.1 Data

The data used to build the model were obtained from the Web Ranking of Universities [14] and the statistics of the International Monetary Fund, published in April 2018. The analysis was made on the first 100 Latin American universities of the Webometrics Ranking.

2.2 Variables

As mentioned above, the dependent variable was the ResearchGate Score and the explanatory variables were: (i) number of postgraduate programs offered by the educational institution (po), (ii) number of teacher's profiles registered in Google Scholar (sp), (iii) number of subscribers to the institutional YouTube channel (YouTube), and (iv) GDP per capita at constant prices adjusted to the purchasing power parity in base dollars 2011 as a control variable of the level of economic development of the origin country of the university, since it is expected that RG score has a positive relationship with quality and impact of the research in its environment (pibppp).

2.3 Model

The log-log model is specified as follows:

$$\mathrm{Lrgj} = \beta 0 + \beta 1\ \mathrm{lpoj} + \beta 2\ \mathrm{lgspj} + \beta 3\ \mathrm{lgyoutubei} + \beta 4 \mathrm{lpibpppi} + \varepsilon_{\mathrm{jt}}. \tag{1}$$

j corresponds to the university; i is the country of origin of the university; lrg is the logarithm of the ResearchGate score; lpo is the logarithm of the number of postgraduate programs offered by the university; lgsp is the logarithm of the number of teacher's profiles registered in Google Scholar; lgyoutube is the logarithm of the number of subscribers to the institutional YouTube channel; lpibppp is the logarithm of GDP per capita at constant prices adjusted to the purchasing power parity in base dollars 2011 of the country where the university is located and ε_{jt} is a random disturbance that is supposed $\varepsilon_{jt} \sim N(0, \sigma^2)$.

3 Results

The results of the model are available below. The proposed model considered all the variables in natural logarithm (Fig. 1).

Source	SS	df	MS
Model	54.4667077	4	13.6166769
Residual	18.0911795	80	.226139744
Total	72.5578872	84	.863784371

Number of obs =	85
F(4, 80) =	60.21
Prob > F =	0.0000
R-squared =	0.7507
Adj R-squared =	0.7382
Root MSE =	.47554

lrg	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lpo	.2328636	.0884934	2.63	0.010	.0567561	.4089711
lgsp	.7892523	.0635073	12.43	0.000	.6628688	.9156359
lgyoutube	-.0914424	.0319105	-2.87	0.005	-.1549463	-.0279386
lpibppp	1.005943	.2797976	3.60	0.001	.4491277	1.562758
_cons	-5.339164	2.858211	-1.87	0.065	-11.02719	.3488574

Fig. 1. Results of the model

In this case, the significant variables are shown. Other variables initially considered as (i) number of undergraduate programs, (ii) number of scientific journals; (iii) number of documents found under the domain of the university; (iv) H-index of the 1st profile of university researcher; (vi) number of members of the institution; (v) Scimago SIR ranking of Higher Education Institutions; (vi) number of tweets published in the institutional account; (vii) number of followers in the Twitter institutional account; (vii) number of "likes" given to the institutional account, were not significant.

To validate the model, the relevant tests were performed for the MCO cases. When performing the Ramsey test, it can be concluded that there was no omission of relevant variables since the null hypothesis can not be rejected. Ramsey RESET test using powers of the fitted values of lrg.

Ho: model has no omitted variables

F (3, 77) = 1.85

Prob > F = 0.1454

The White's test shows that the variance of random perturbations, conditional on the values of the regressors, are constant. Since the null hypothesis of homoskedasticity can not be rejected, as shown below (Fig. 2):

White's test for Ho: homoscedasticity against

Ha: unrestricted heteroskedasticity

chi2(14) = 10.59

Prob > chi2 = 0.7180

```
---------------------------------------------------
Source | chi2 df p
--------------------+----------------------------
Heteroskedasticity | 10.59 14 0.7180
Skewness | 5.45 4 0.2441
Kurtosis | 3.18 1 0.0745
--------------------+----------------------------
Total | 19.22 19 0.4428
---------------------------------------------------
```

Fig. 2. Cameron & Trivedi's decomposition of IM-test

To test for normality of the errors, the tests of normality of kurtosis and pointing, and Shapiro-Wilk were made, concluding that the null hypothesis of normality to a significance level of 5% can not be rejected (Fig. 3).

```
------- joint ------
Variable | Obs Pr (Skewness) Pr (Kurtosis) adj chi2(2) Prob>chi2
-------------+------------------------------------------------------------
Residual | 85 0.1073 0.0939 5.27 0.0717
swilk residual
Shapiro-Wilk W test for normal data
Variable | Obs W V z Prob>z
-------------+-------------------------------------------------
Residual | 85 0.97859 1,545 0,956 0.16950
```

Fig. 3. Skewness/Kurtosis tests for Normality

Finally, the variance inflation factor (VIF) was calculated to prove that there was no multicollinearity in the independent variables, concluding that there was no multicollinearity since the value was less than 3 (Fig. 4).

Variable	VIF	1/VIF
lgsp	1.51	0.660182
lpo	1.27	0.785697
lpibppp	1.18	0.849824
lgyoutube	1.05	0.950065
Mean VIF	1.25	

Fig. 4. Variance inflation factor (VIF)

Since the model does not violate any of the MCO assumptions, no transformation is necessary. The coefficient of determination is 73.82%. The significant variables in the model were the number of postgraduate programs offered by the university; the number of teacher's profiles registered in Google Scholar; the number of subscribers to the institutional YouTube channel and the GDP per capita of the country where the university is located.

Therefore, the equation of the model corresponds to:

$$\begin{aligned}\mathrm{Lrgj} = & -5.339164 + 0.2328636 * \mathrm{lpoj} + 0.7892523 * \mathrm{lgspj} + -0.914424 * \mathrm{lgyoutubei} \\ & + 1.005943 * \mathrm{pibpppi} + \varepsilon_{jt}.\end{aligned} \tag{2}$$

The effects of each of the variables are explained below.

- An increase of 1% in the number of postgraduate programs offered by the university increases the ResearchGate score of the institution by 0.23%.
- An increase of 1% in the number of teacher's profiles registered in Google Scholar increases the ResearchGate score of the institution by 0.79%.
- An increase of 1% in the number of subscribers to the institutional YouTube channel reduces the ResearchGate score of the institution by 0.9144%.
- An increase of 1% in GDP per capita at constant prices adjusted to the purchasing power parity in base dollar 2011 of the country where the university is located increases the ResearchGate score of the institution by 1.006%.

4 Conclusions

The model of standard errors corrected for the model has a determination coefficient of 73.82%. The only variable that presented a negative relationship with the dependent variable was the number of subscribers to the institutional YouTube channel, that is, the popularity on YouTube of the institution has an inverse effect on the RG score, what can be explained since YouTube corresponds to a social network and not to an academic one.

Variables such as (i) number of undergraduate programs, (ii) number of scientific journals; (iii) number of documents found under the university domain; (iv) H-index of the 1st profile of researcher at the university; (vi) number of members of the institution; (v) SIR Scimago ranking of Higher Education Institutions; (vi) number of tweets published in the institutional account; (vii) number of followers in the Twitter institutional account; (vii) number of "likes" given to the institutional count, were not significant.

Considering the above, some of the recommendations for the institutions to increase their position in the RG score are: (i) increase the number of teachers with active profiles in academic social networks, and (ii) increase the number of postgraduate programs and postgraduate students.

For future researches, this analysis could be carried out on other rankings such as Shanghai, QS World University Ranking, and SCimago Institutions Rankings SIR. Those results could allow to provide recommendations to the academic authorities of higher education institutions to increase their visibility.

References

1. ReseachGate the network for science and research (2018). https://solutions.researchgate.net/recruiting/?utm_source=researchgate&utm_medium=community-loggedout&utm_campaign=indextop
2. Orduña-Malea, E., Martín-Martín, A., López-Cózar, E.D.: ResearchGate como fuente de evaluación científica: desvelando sus aplicaciones bibliométricas. El profesional de la información (EPI) **25**(2), 303–310 (2016)
3. Yu, M.C., Wu, Y.C.J., Alhalabi, W., Kao, H.Y., Wu, W.H.: ResearchGate: an effective altmetric indicator for active researchers? Comput. Hum. Behav. **55**, 1001–1006 (2016)
4. Yan, W., Zhang, Y.: Research universities on the ResearchGate social networking site: an examination of institutional differences, research activity level, and social networks formed. J. Inform. **12**(1), 385–400 (2018)
5. Thelwall, M., Kousha, K.: ResearchGate: disseminating, communicating, and measuring scholarship? J. Assoc. Inf. Sci. Technol. **66**(5), 876–889 (2015)
6. Martín-Martín, A., Orduña-Malea, E., Ayllón, J.M., López-Cózar, E.D.: The counting house: Measuring those who count. Presence of bibliometrics, scientometrics, informetrics, webometrics and altmetrics in the Google Scholar citations, Researcherid, ResearchGate, Mendeley & Twitter (2016). arXiv preprint arXiv:1602.02412
7. Torres-Samuel, M., Vásquez, C.L., Viloria, A., Varela, N., Hernández-Fernandez, L., Portillo-Medina, R.: Analysis of patterns in the university world rankings webometrics, Shanghai, QS and SIR-SCimago: case Latin America. In: Tan, Y., Shi, Y., Tang, Q. (eds.) DMBD 2018. LNCS, vol. 10943, pp. 188–199. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93803-5_18
8. Torres-Samuel, M., et al.: Efficiency analysis of the visibility of Latin American universities and their impact on the ranking web. In: Tan, Y., Shi, Y., Tang, Q. (eds.) DMBD 2018. LNCS, vol. 10943, pp. 235–243. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93803-5_22
9. Vásquez, C., et al.: Cluster of the Latin American universities Top100 according to webometrics 2017. In: Tan, Y., Shi, Y., Tang, Q. (eds.) DMBD 2018. LNCS, vol. 10943, pp. 276–283. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93803-5_26
10. Torres-Samuel, M., Vásquez, C., Viloria, A., Lis-Gutiérrez, J.P., Borrero, T.C., Varela, N.: Web visibility profiles of Top100 Latin American universities. In: Tan, Y., Shi, Y., Tang, Q. (eds.) DMBD 2018. LNCS, vol. 10943, pp. 254–262. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93803-5_24
11. Ramanayaka, K.H., Chen, X., Shi, B.: Application of webometrics techniques for measuring and evaluating visibility of university library websites in Sri Lanka. J. Univ. Libr. Assoc. Sri Lanka **21**(1), 1–17 (2018). https://doi.org/10.4038/jula.v21i1.7908
12. Jati, H.: University webometrics ranking using multicriteria decision analysis: entropy and TOPSIS method. Soc. Sci. **13**(3), 763–765 (2018). http://docsdrive.com/pdfs/medwelljournals/sscience/2018/763-765.pdf
13. Manoj, M.: Webometrics as a tool for measuring popularity of websites: an analysis of websites of IISERs in India. Int. J. Sci. Res. Sci. Technol. **4**(2), 1472–1476 (2018)
14. Cybermetrics Lab Webometrics Ranking of World Universities (2018). http://www.webometrics.info/en/About_Us