University Website Ranking: Webometrics v. GPR

György Losonczi

Szent István University
gLosonczi@gmail.com
+36-30-9344-981
Research fields since 2011: Online Marketing, Decision Making, Datamining, Website Evaluation of different segments, e.g. university, national museums, memorial institutions, and circus.

This study addresses the issue of the Internet rankings of the universities websites. Why is it important for the institutions to have a competitive website and ranked in good positions? Because websites are the most important form of their online appearance, as a reflection of the style, the activity and the reputation of the particular institution. The study evaluates the two most known website rankings: WEBOMETRICS and Google PageRank.

Keywords: Webometrics, Google PageRank, website, higher education, competitiveness, online marketing

Introduction

Why is it important for the institutions to have a well-ranked website, other words competitive website? Because websites are the most important form of their online appearance, as a reflection of the style, the activity and the reputation of the particular institution. (Spencer – Ruwoldt, 2004) Over the past six years, I was evaluating the competitiveness of the websites of higher education institutions and it is still a part of my Ph.D. research. This study addresses the issue of the online rankings of the university websites using the well known WEBOMETRICS and GOOGLE PageRank. Both are using "weights" in their defined methodology. The goals of this research: “Is there a relationship among WEBOMETRICS and GOOGLE PageRank?“

2. Methodology, Ratings Overview

2.1 Methodology, Dataset

The data assets contain more than ten thousand institutional websites and their appearance of rating values, which refers to the year 2014. More recent data for WEBOMETRICS are still publicly available, while Google PageRank is no longer available for the public since 2014. Therefore my analysis was based on the rating values of the year of 2014. The data assets also contain the other four parts (rankings) of WEBOMETRICS: “Presence Rank”, “Impact”, “Openness”, “Excellence Rank”.

	Webometrics World Rank (2014)	University	Domain	Presence Rank*	Impact Rank*	Opennes s Rank*	Excellenc e Rank*	Google PageRan k (2014)
1	1	Harvard University	edu	3	3	8	1	8
4	4	Cornell University	edu	13	5	2	21	8
5	5	Columbia University New York	edu	52	8	35	10	8
6	6	University of California Berkeley	edu	2090	1	42	15	8
7	7	University of Pennsylvania	edu	14	12	93	9	8
8	8	University of California Los Angeles UCLA	edu	43	16	55	4	8
14	14	(2) Johns Hopkins University	edu	103	59	17	2	8
15	15	Texas A&M University	edu	23	22	25	84	8
16	16	Princeton University	edu	19	20	58	72	8
18	18	University of Florida	edu	88	31	21	38	8
19	19	Purdue University	edu	55	19	44	74	8
23	23	Carnegie Mellon University	edu	54	30	9	141	8
26	26	(1) University of Maryland	edu	50	35	38	67	8
28	27	University of Southern California	edu	232	25	56	47	8
29	29	(2) Universidade de São Paulo USP	br	33	54	10	78	8
30	30	University of British Columbia	ca	90	46	50	26	8
31	31	University of Minnesota	edu	3918	6	28	23	8

1. Figure The dataset in Excel

Source: Created by the author using Excel

The dataset contains 11 891 records, Figure 1 demonstrates a part of it. In my paper I used the following important methods to test the correlation between the two rankings:

- Scatter diagram method,
- Box-Plot,
- Spearman's rank correlation coefficient,
- Kendall rank correlation.

For the interpretation of a correlation coefficient I used Cohen's guidelines (Cohen, 1992):

Correlation coefficient value	Association
-0.3 to +0.3	Weak
-0.5 to -0.3 or 0.3 to 0.5	Moderate
-0.9 to -0.5 or 0.5 to 0.9	Strong
-1.0 to -0.9 or 0.9 to 1.0	Very strong

1. Table Cohen's guideline for the interpretation of a correlation coefficient

Source: (Marshall, n.d.)

2.2 Short Introduce of WEBOMETRICS

The Webometrics Ranking of World Universities (since 1997), also known as Ranking Web of Universities is published by the Cybermetrics Lab, a research group of the Spanish National Research Council (CSIC). The ranking system based on a composite indicator that takes into account the following aspects:

- Size (S) aspect: „Presence Rank" - Number of pages recovered from search engines (10%);
- Visibility (V) aspect: „Impact Rank" – The total number of unique external links received (50%);
- Rich Files (R) aspect:„Openness Rank" – academic and publication activities (10%);
- Scholar (Sc) aspect: „Excellence Rank" – Scimago (30%).

By now The Webometrics Ranking of World Universities updated their methodology. (WEBOMETRICS.INFO, 2016) Figure 2 shows the Comparison of the main World Universities's Rankings.

CRITERIA	WR (webometrics)		ARWU (Shanghai)	
Univ's Analyzed	15000		3000	
Univ's Ranked	5000+		500	
Quality of Education			Alumni Nobel&Field	10%
Internazionalization				
Size	Web Size	20%	Size of Institution	10%
Research Output	Rich Files	15%	Nature & Science	20%
	(Google) Scholar	15%	SCI & SSCI	20%
Impact	(Link) Visibility	50%	Highly Cited Res'ers	20%
Prestige			Staff Nobel&Field	20%

2. Figure Comparison of the main world universities' rankings

Source: http://www.webometrics.info/en/Objetives)

2.3 Short Introduce of Google PageRank

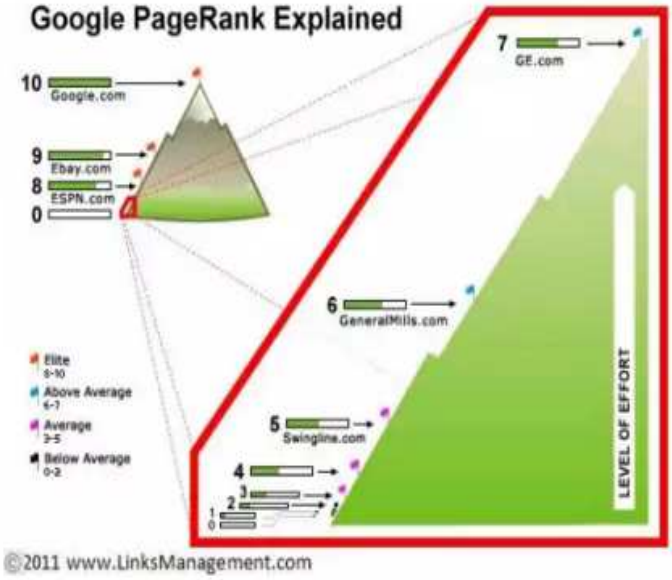
PageRank was named after one of the founders of Google company: Larry Page. PageRank is a unique measuring "gauge" of website pages, especially their importance on the Internet.

Google says the Google PageRank is a complex algorithm of link analysis:

„Today Google's algorithms rely on more than 200 unique signals or "clues" that make it possible to guess what you might really be looking for. These signals include things like the terms on websites, the freshness of content, your region and PageRank." (Brin – Page, 2016)

PageRank is a number between 0 and 10, a page rank zero or no page rank means that the website is not listed, rating value 10 is the highest rating on the PageRank scale. Figure 3 presents the relationship between PageRank and Backlink Page.

Google PR of the Backlink Page	PR Weight of the Backlink Page
PR0	1
PR1	5
PR2	25
PR3	125
PR4	625
PR5	3,125
PR6	15,625
PR7	78,125
PR8	390,625
PR9	1,953,125
PR10	9,765,625



3. Figure Google PageRank

Source: http://repcapitalmedia.com/how-to-increase-your-google-plus-pagerank

3. Result

The research has been conducted among 12 000 university websites (173 countries) and specially focused on their first landing page. All of them were evaluated in the year 2014 and saved to the database. Table 2 presents the frequency of the PageRank values.

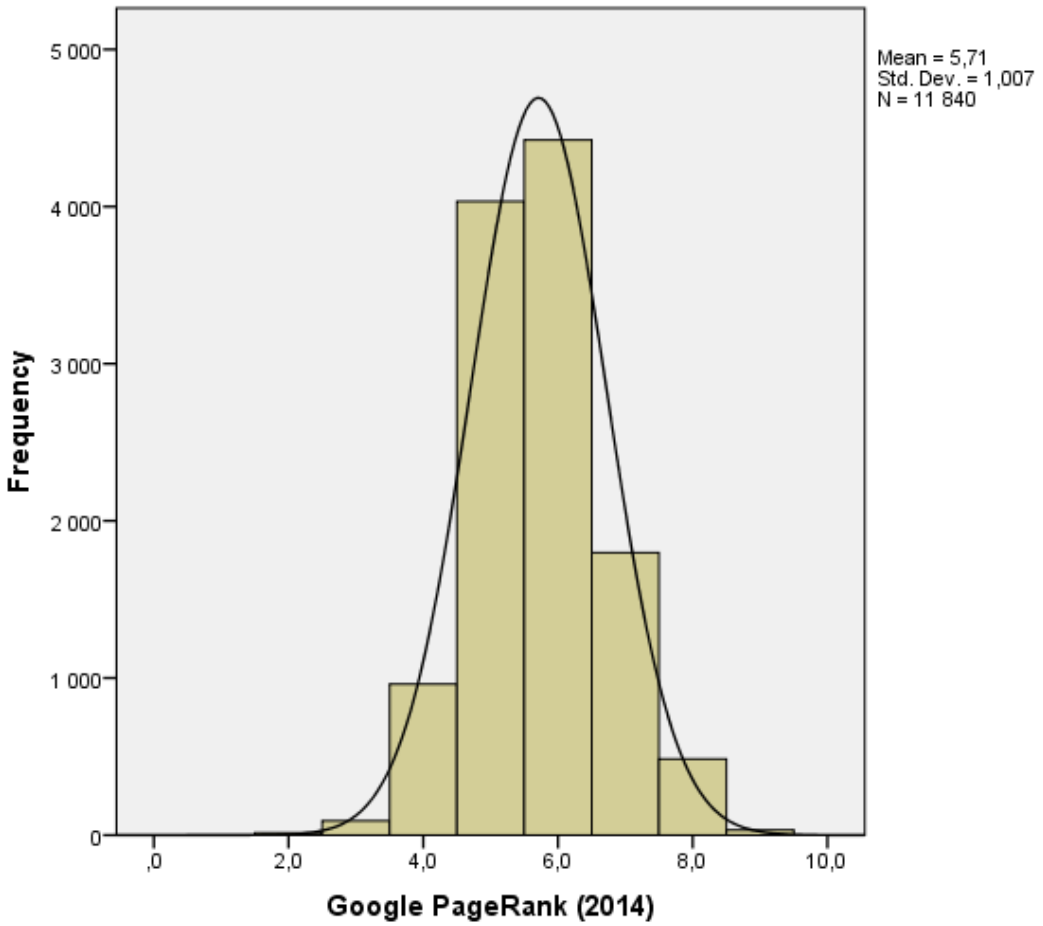
The dataset contains 11 891 records, but for the analysing, I did not use a GPR 0 value (websites without ranking) and there is no website with the ranked value GPR 10; the total record is 11 840. The mean of the PageRank of the WEBOMETRICS ranked universities is 5.71; the standard deviation is 1.0065. The median is 6.00 and the mode is also 6.00; the frequency distribution of scores is quite symmetrical as we can see in Figure 4.

Google PageRank (2014)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,0	1	,0	,0	,0
	2,0	15	,1	,1	,1
	3,0	91	,8	,8	,9
	4,0	961	8,1	8,1	9,0
	5,0	4033	34,1	34,1	43,1
	6,0	4424	37,4	37,4	80,4
	7,0	1798	15,2	15,2	95,6
	8,0	485	4,1	4,1	99,7
	9,0	32	,3	,3	100,0
	Total	11840	100,0	100,0	

2. Table Frequency of the universities PageRanks in the WEBOMETRICS ranking

Source: Created by the author using SPSS



4. Figure Histogram of the universities PageRanks in the WEBOMETRICS ranking

Source: Created by the author using SPSS

The histogram shape is not multi-modal; verifies a close symmetry, lighly asymmetric as the skewness value is 0.165, lighly "skewed right" distribution. Kurtosis value is 2.58 (SPSS result, Table 3) quite high as in SPSS a normal distribution has kurtosis 0 and leptokurtic as the high frequencies are only a small part of the curve.

Google PageRank (2014)

N	Valid	11840
	Missing	0
Mean		5,710
Median		6,000
Mode		6,0
Std. Deviation		1,0065
Skewness		,165
Std. Error of Skewness		,023
Kurtosis		,258
Std. Error of Kurtosis		,045
Range		8,0
Sum		67611,0
Percentiles	25	5,000
	50	6,000
	75	6,000

3. Table Statistics of the universities PageRanks in the WEBOMETRICS ranking (1)
Source: Created by the author using SPSS

Table 4 shows the statistics for each PageRank value category and Table 5 demonstrate another aspect of grouping the results and it's weight in the dataset.

Webometrics World Rank (2014)

Google PageRank (2014)	N	% of Total N	Mean	Median	Variance	Skewness	Kurtosis
1,0	1	0,0%	5940,000	5940,000	.	.	.
2,0	15	0,1%	8235,533	8774,000	9020698,410	-,858	-,112
3,0	91	0,8%	9927,802	10746,000	5373592,649	-2,141	4,967
4,0	961	8,1%	9527,777	9925,000	3851330,098	-1,063	,788
5,0	4033	34,1%	7470,741	7592,000	7270294,485	-,198	-,932
6,0	4424	37,4%	5555,685	5181,500	9419340,257	,322	-1,000
7,0	1798	15,2%	2926,765	1928,000	7350171,707	1,316	1,069
8,0	485	4,1%	1473,289	509,000	5574680,755	2,484	5,794
9,0	32	0,3%	401,813	195,000	441661,835	3,105	9,950
Total	11840	100,0%	5987,050	5977,000	12007722,64	,004	-1,201

4. Table Statistics of the universities PageRanks in the WEBOMETRICS ranking (2)
Source: Created by the author using SPSS

	Google PageRank (2014)	Mean	N	Variance	% of Total N	Median	Kurtosis	Skewness
	0	7065,176	51	10542600,708	,4%	7363,000	-,978	-,422
9.0%	1	5940,000	1		,0%	5940,000		
	2	8235,533	15	9020698,410	,1%	8774,000	-,112	-,858
	3	9927,802	91	5373592,649	,8%	10746,000	4,967	-2,141
	4	9527,777	961	3851330,098	8,1%	9925,000	,788	-1,063
71.1%	5	7470,741	4033	7270294,485	33,9%	7592,000	-,932	-,198
	6	5555,685	4424	9419340,257	37,2%	5181,500	-1,000	,322
15.1%	7	2926,765	1798	7350171,707	15,1%	1928,000	1,069	1,316
4.4%	8	1473,289	485	5574680,755	4,1%	509,000	5,794	2,484
	9	401,813	32	441661,835	,3%	195,000	9,950	3,105
	Total	5991,674	11891	12005515,924	100,0%	5986,000	-1,201	,002

Webometrics World Rank (2014)

	Google PageRank (2014)	N	% of Total N	Mean	Median	Variance	Skewness	Kurtosis
	1,0	1	0,0%	5940,000	5940,000	.	.	.
9.0%	2,0	15	0,1%	8235,533	8774,000	9020698,410	-,858	-,112
	3,0	91	0,8%	9927,802	10746,000	5373592,649	-2,141	4,967
	4,0	961	8,1%	9527,777	9925,000	3851330,098	-1,063	,788
71.5%	5,0	4033	34,1%	7470,741	7592,000	7270294,485	-,198	-,932
	6,0	4424	37,4%	5555,685	5181,500	9419340,257	,322	-1,000
15.2%	7,0	1798	15,2%	2926,765	1928,000	7350171,707	1,316	1,069
4.4%	8,0	485	4,1%	1473,289	509,000	5574680,755	2,484	5,794
	9,0	32	0,3%	401,813	195,000	441661,835	3,105	9,950
	Total	11840	100,0%	5987,050	5977,000	12007722,64	,004	-1,201

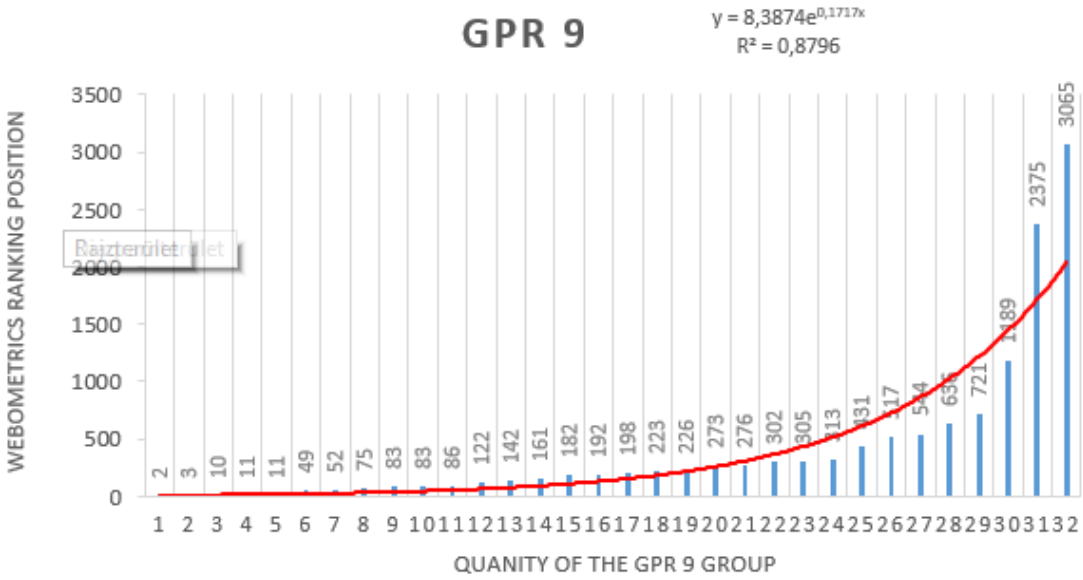
5. Table Statistics of the universities PageRanks in the WEBOMETRICS ranking (3)
Source: Created by the author using SPSS

Figure 5 demonstrate the trends of each value group of PageRank created in Excel using exponential trend option. The horizontal axis is presenting the frequencies of each ranking positions in WEBOMETRICS, the vertical axis presents the WEBOMETRICS positions.

5. Figure Trends of the universities PageRanks in the WEBOMETRICS ranking

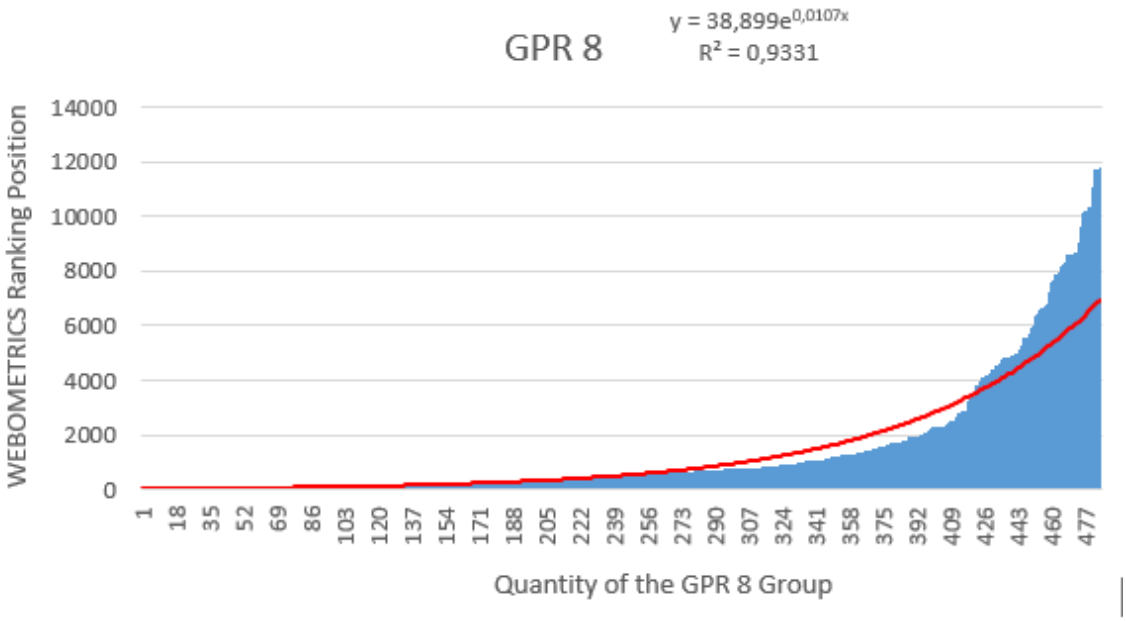
Source: Created by the author using SPSS

Figure 6 and 7 reveals that university websites with PageRank value 8 and 9 have higher ranking positions in WEBOMETRICS ranking.



6. Figure PageRank 9 and WEBOMETRICS positions

Source: Created by the author using SPSS

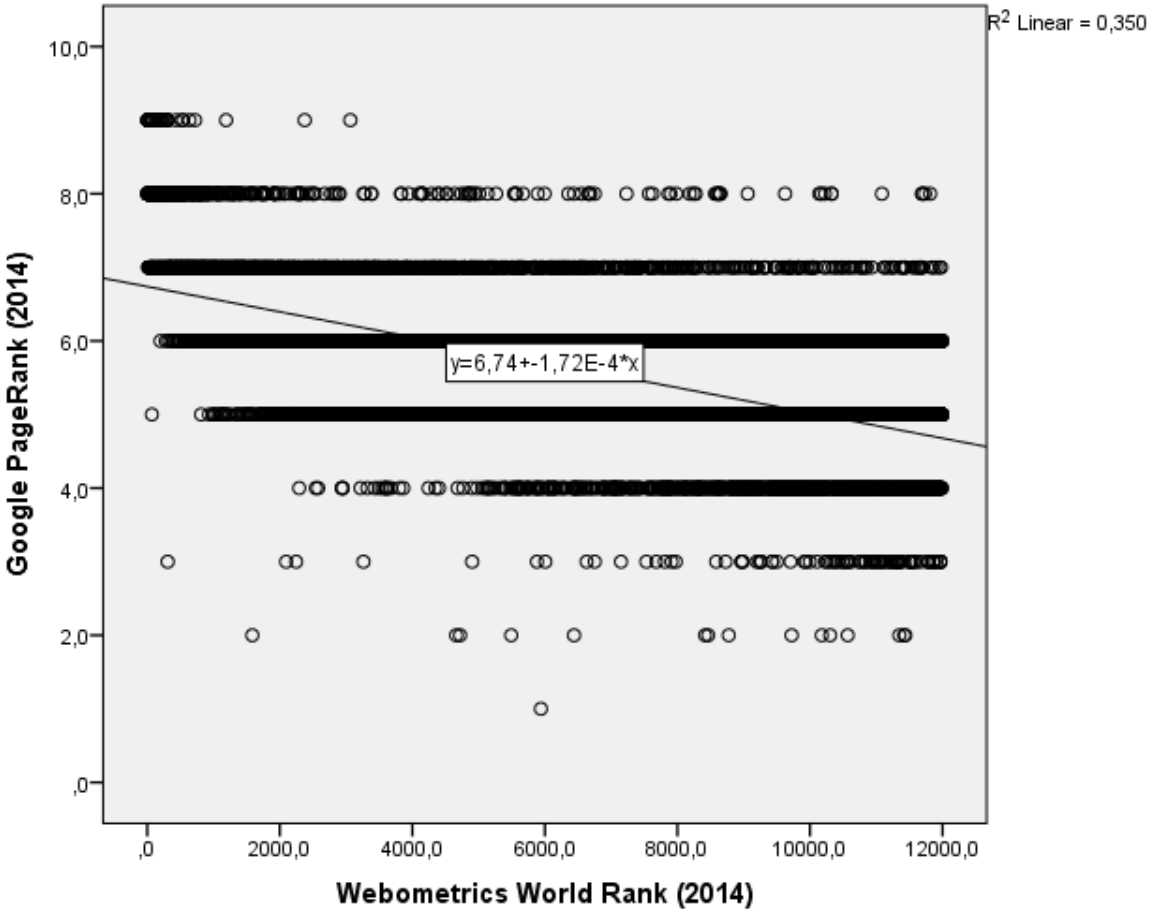


7. Figure PageRank 8 and WEBOMETRICS positions

Source: Created by the author using SPSS

Scatter diagram method is a simple tool for studying the correlation between two variables using a special type of dot chart: the Scatterplot. By applying Scatterplot we can get an idea about the direction of correlation, and also whether it is strong or weak. Figure 8) reveals PageRank value 9, 4 and 3 concentrate in the higher or lower level of ranking positions in WEBOMETRICS.

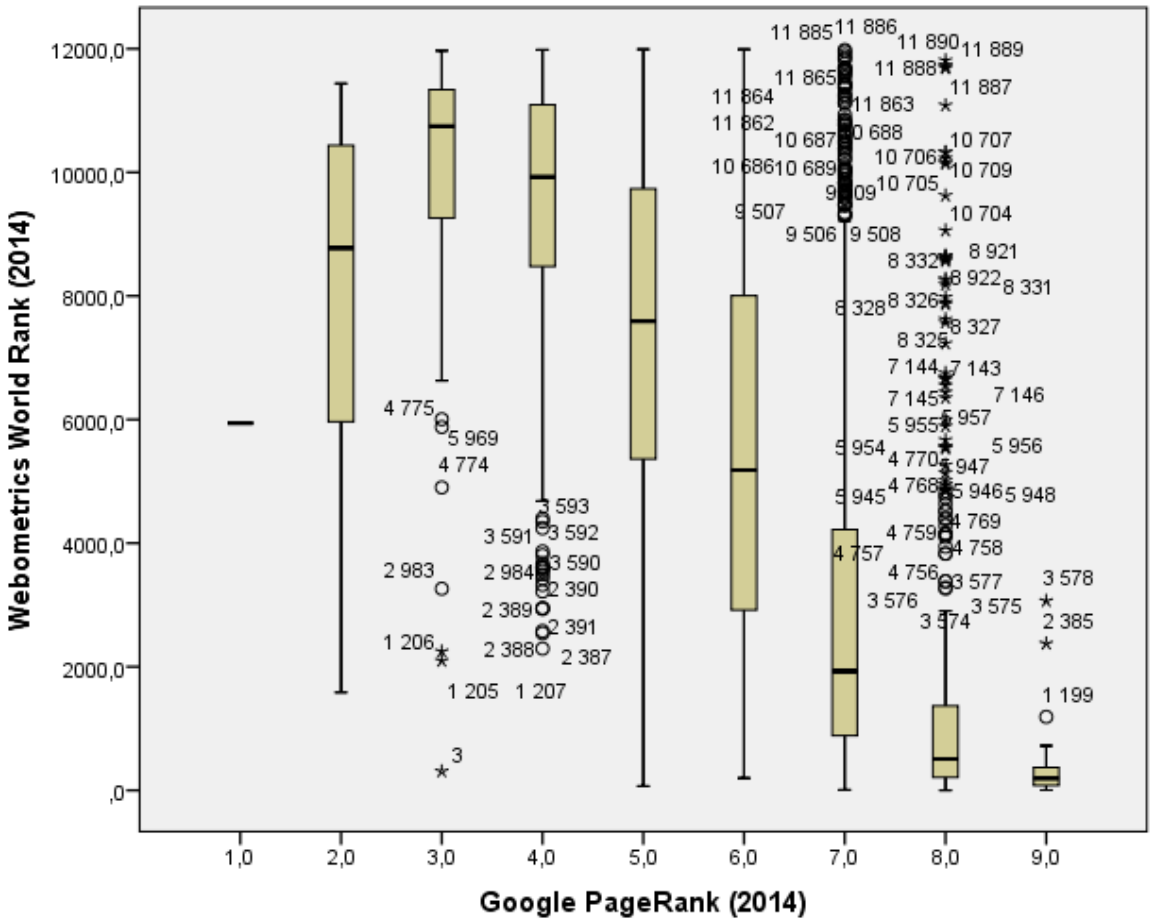
On the scatterplot and box plot shows a moderate negative monotonic relationship between the two variables.



8. Figure Scatterplot of GPR - WEBOMETRICS

Source: Created by the author using SPSS

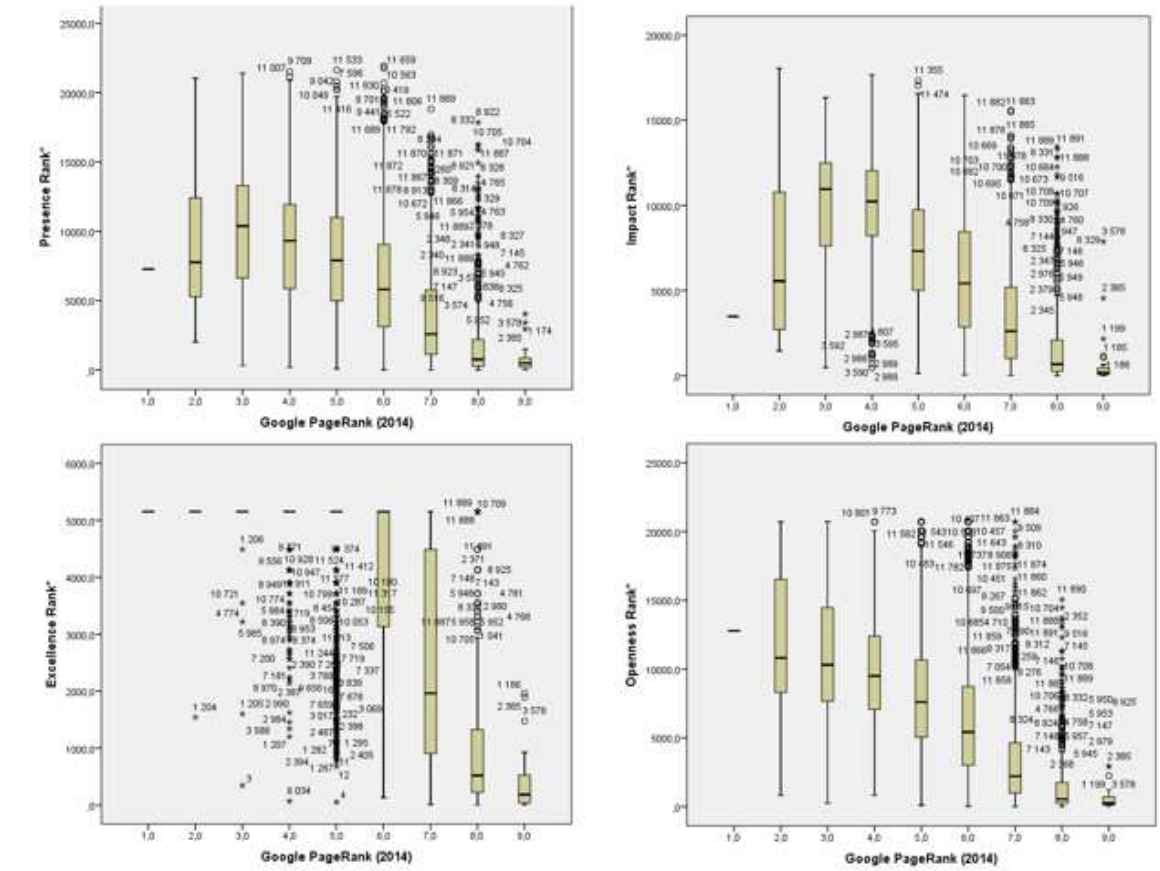
Beside the histogram and scatterplot we can use a box plot to graphically summarize the data set, also called a "five-number summary" of the distribution. Similarly to the Excel diagrams, it demonstrates the same conclusion on Figure 9. Higher ranked WEBOMETRICS universities has better PageRank rating especially at GPR 8 and 9.



9. Figure Box Plot of WEBOMETRICS positions

Source: Created by the author using SPSS

The ranking system based on four composite indicators: „Presence Rank" (Size aspect); „Impact Rank" (Visibility aspect); „Openness Rank" (Rich Files aspect) and „Excellence Rank" (Scholar aspect). The following Figure shows the four indicator"s Box-Plot results. As we see the Box-Plot of „Presence Rank" (top left), „Impact Rank" (top right); „Openness Rank" (bottom right) are quite similar, but „Excellence Rank" (bottom left) differ from the three other. The first three indicators are web presence orientated and have a stronger relationship with the WEBOMETRICS ranking (see later Spearman rank correlation results), but „Excellence Rank" scholar orientated and less correlated. More detailed figures can found in the Appendix.



10. Figure Box Plot of four indicators of WEBOMETRICS

Source: Created by the author using SPSS

I used Spearman rank correlation to test the association between the two ranked variables. It is not very sensitive to outliers, which are observations within your data that do not follow the usual pattern. The rank correlation coefficient measures the degree of similarity between the two rankings and can be used to assess the significance of the relation between them.

Table 6 presents the output of SPSS result, the correlation coefficient is -0,587 shows a strong stochastic correlation. The p-value is less than .01, then we have evidence of a statistically significant bivariate association between the two ordinal variables.

			Google PageRank (2014)	Webometrics World Rank (2014)
Spearman's rho	Google PageRank (2014)	Correlation Coefficient	1,000	-,587**
		Sig. (2-tailed)	.	,000
		N	11840	11840
	Webometrics World Rank (2014)	Correlation Coefficient	-,587**	1,000
		Sig. (2-tailed)	,000	.
		N	11840	11840
	Presence Rank*	Correlation Coefficient	-,442**	,746**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Impact Rank*	Correlation Coefficient	-,500**	,895**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Openness Rank*	Correlation Coefficient	-,507**	,790**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Excellence Rank*	Correlation Coefficient	-,537**	,691**
		Sig. (2-tailed)	,000	,000
		N	11840	11840

**. Correlation is significant at the 0.01 level (2-tailed).

6. Table Spearman's rho result
Source: Created by the author using SPSS

Table 6 reveals that „Presence Rank", „Impact Rank", „Openness Rank" have a stronger relationship with the WEBOMETRICS but „Excellence Rank" is less correlated. Also remarkable result that „Excellence Rank" has a stronger relationship with GOOGLE PageRank.

Beside Spearman rank correlation I also used Kendall rank correlation to measure the ordinal association between two measured quantities. Contrary to the Spearman correlation, the Kendall correlation is not affected by how far from each other ranks are. Only appropriate for discrete variables as it analyses only by whether the ranks between observations are equal or not. Table 7 presents the output of SPSS result, the correlation coefficient is -0,468 shows a moderate stochastic correlation. Kendall's tau usually smaller values than Spearman's rho correlation and in my results that were exactly happened.

			Google PageRank (2014)	Webometrics World Rank (2014)
Kendall's tau_b	Google PageRank (2014)	Correlation Coefficient	1,000	-,468**
		Sig. (2-tailed)	.	,000
		N	11840	11840
	Webometrics World Rank (2014)	Correlation Coefficient	-,468**	1,000
		Sig. (2-tailed)	,000	.
		N	11840	11840
	Presence Rank*	Correlation Coefficient	-,343**	,562**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Impact Rank*	Correlation Coefficient	-,393**	,719**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Openness Rank*	Correlation Coefficient	-,397**	,600**
		Sig. (2-tailed)	,000	,000
		N	11840	11840
	Excellence Rank*	Correlation Coefficient	-,464**	,556**
		Sig. (2-tailed)	,000	,000
		N	11840	11840

**. Correlation is significant at the 0.01 level (2-tailed).

7. Table Kendall's tau result
Source: Created by the author using SPSS

4 Conclusion

The study addresses the issue of the online rankings of the university websites WEBOMETRICS and GOOGLE PageRank and researching relationship among the two rankings.

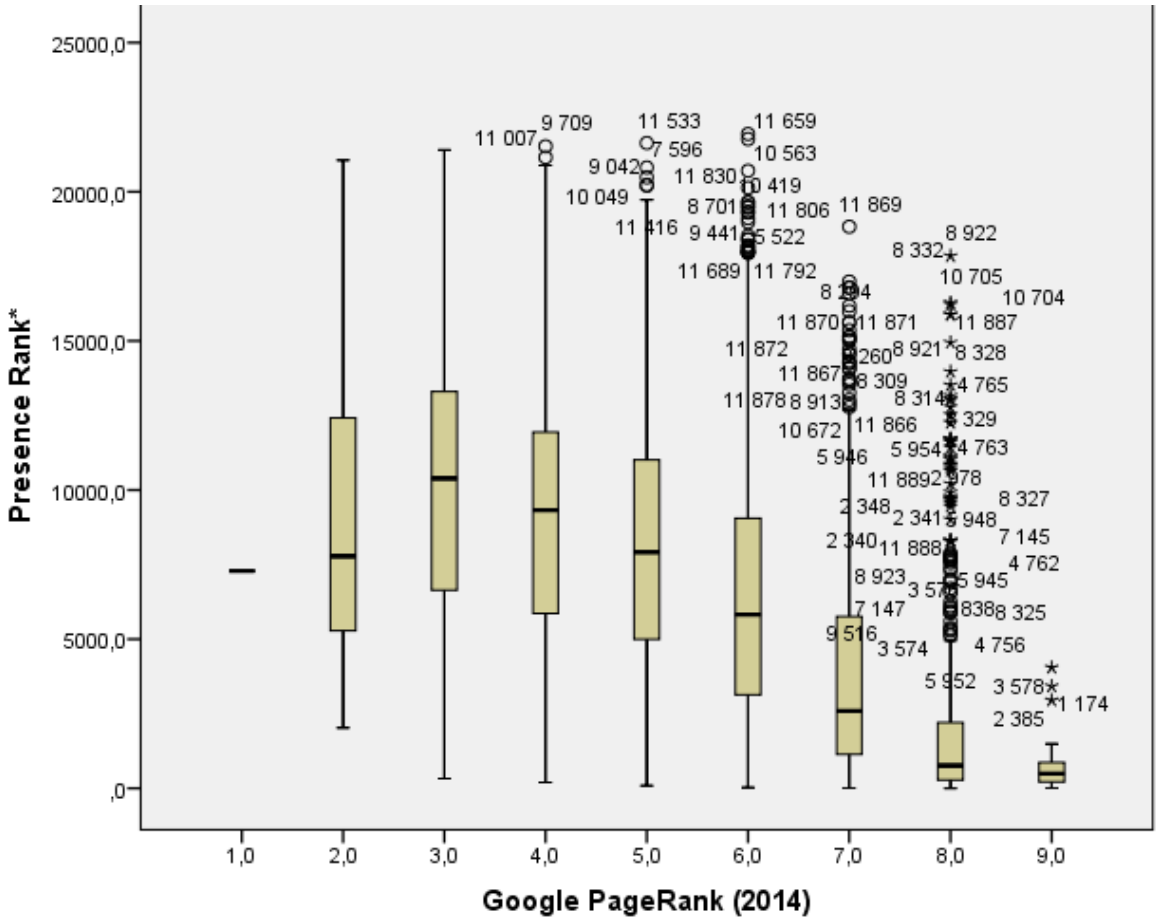
During the evaluation, it is revealed that university websites with PageRank value 8 and 9 have higher ranking positions in WEBOMETRICS ranking. Spearman rank correlation result was strong stochastic correlation while Kendall rank correlation was a moderate stochastic correlation among the two rankings. Also remarkable result that „Excellence Rank" has a stronger relationship with GOOGLE PageRank than the other presence orientated indicators of WEBOMETRICS.

Final conclusion: for more facts the research needs to involve website metatag data to analyze the deeper relationship among the rankings.

References

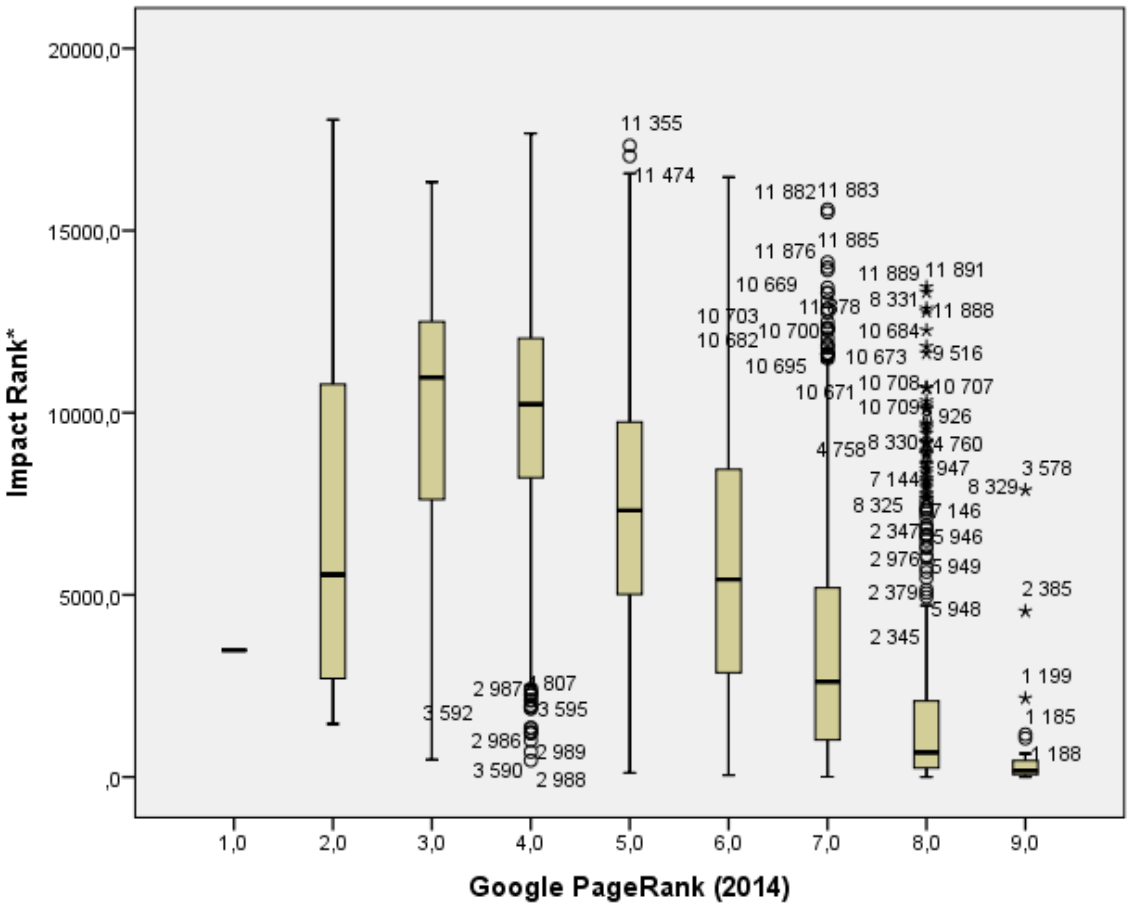
1. Brin, S.; Page, L., 2016, *The Anatomy of a Large-Scale Hypertextual Web Search Engine* (Online). http://infolab.stanford.edu/~backrub/google.html. (downloaded: 10/09/16 23:09)
2. Cohen, L., 1992. *Power Primer*. Psychological Bulletin, 112(1) 155-159
3. GOOGLE, 2016, *Algorithms – Inside Search* (Online). https://www.google.com/insidesearch/howsearchworks/algorithms.html. (downloaded: 10/09/16 23:09)
4. Marshall, E., n.d. *Scatterplots and correlation in SPSS* (Online). http://www.sheffield.ac.uk/polopoly_fs/1.531428!/file/MASHScatterplot_correlation_SPSS.pdf (downloaded: 31/03/17 10:00)
5. Spencer, C., Ruwoldt, M. L., 2004. *Navigation and content on university home pages*, (Online), http://repository.unimelb.edu.au/10187/1141 (downloaded: 25/5/10 2010 9:14)
6. WEBOMETRICS.INFO, 2016, *Updated methodology* (Online), http://www.webometrics.info/en/node/19 (downloaded: 10/09/16 23:05)

Appendix



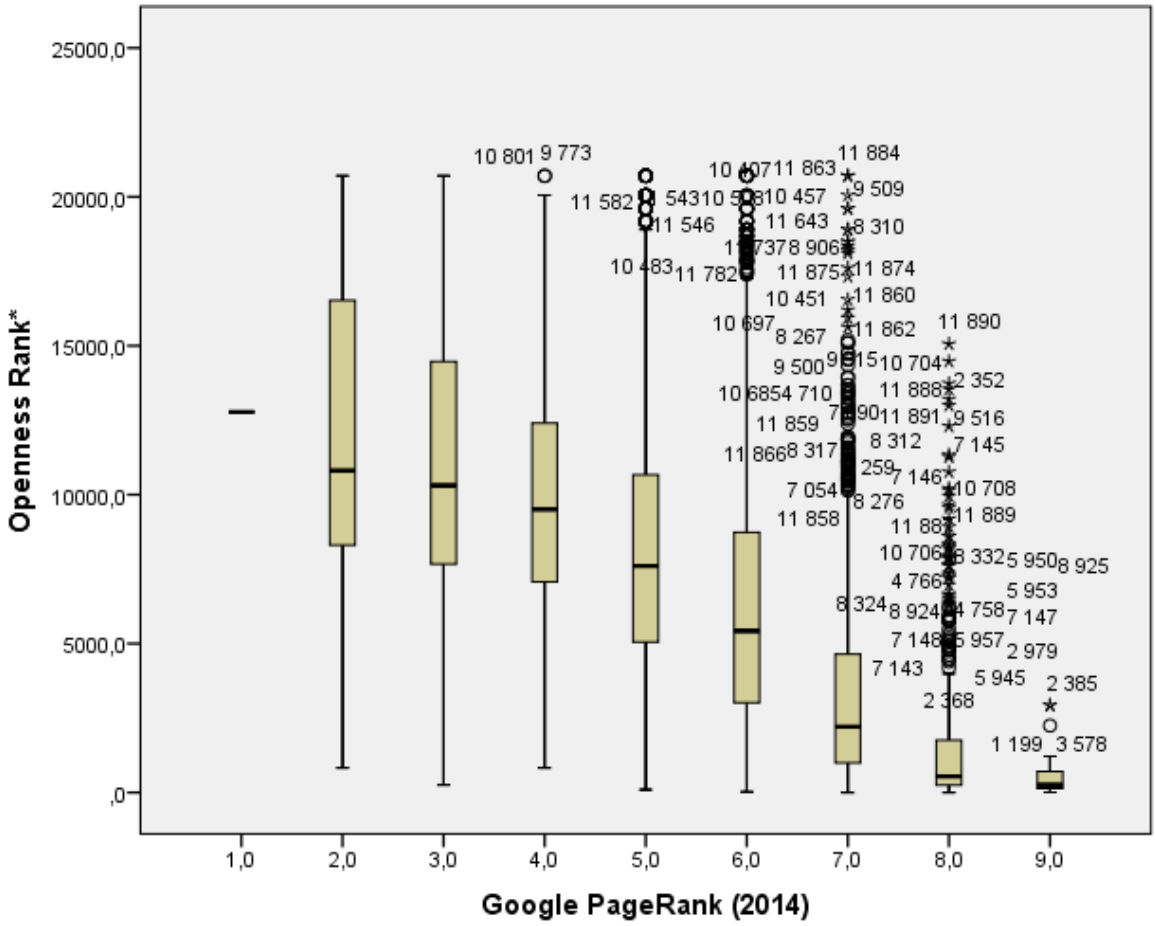
11. Figure Box Plot of Presence Rank of WEBOMETRICS

Source: Created by the author using SPSS



12. Figure 7 Box Plot of Impact Rank of WEBOMETRICS

Source: Created by the author using SPSS



13. Figure Box Plot of Openness of WEBOMETRICS
Source: Created by the author using SPSS