

Study of the Relationship Between World University Rankings and Count of Links to Them: Based on Shanghai University Ranking and Times Higher Education

Abdolrasoul Jowkar¹ & Reza Rajabali Beglou²

29th of June 2008

Rankings have always attracted the attention of many towards themselves. People have always liked to know for example what, who, which city or country have the highest or the lowest rankings from different aspects, so that they can examine, compare and improve those aspects according to the result of rankings. Meanwhile, with the increase of attention to universities and post-graduate institutions, university ranking has gained larger importance. University rankings-not considering their problems- is very important in attracting the attention of those who are after finding and choosing a suitable university for their further education. On the other hand, connection with websites has also drawn the attention of the web researchers to itself. Web evaluation studies, as part of web studies, is possible through the web links, which have been founded by Ingoursen. Web researchers have understood that links are potentially capable of revealing scientific and evaluative relationships of the web. Correlation between the geographical distances of universities and researches performed in them, and a survey of the relation of web links and research products are kinds of scientific relationships which have been used in the scientific products of these organizations and educational centers. A survey of "valid links", that is, links which reaffirm the existence of a relation between websites, shows that there is improvement in the web researchers, that of web links. However, although because of the differences in the types of media we can not exactly conform to former theories in events such as journal documentation and research products with a calculation of links and rate of university research products, such great differences cannot decrease the importance of surveying the relation of the rate of links and research products. A survey of the intentions behind creating links by the producers of web pages has almost decreased the complications of this issue.

Statement of the problem:

According to Bjornborn's definition "webometrics is the study of the quantitative aspects of creating and using information sources, web structures and technologies which is possible through drawing informometrics and bibliometrics approaches". The term and concept of the web influence factor was first used by Gairin in a Spanish journal in 1997 and till Ingorsun's time did not attract much attention. The percent research, with the background of accomplished studies, intends to study an aspect of webometrics, namely, survey the relation between university rankings and the rate of their links. Most of the studies of webometrics have been performed on electronic journals and university websites. "It seems that since documentation has been created and used in scientific environments, it can be used as an evidence for understanding scientific relations" (Koosha, 2003). The World Wide Web Consortium does not present a rule for creating links but only has the role of "guiding the world wide web by developing common protocols in order to perform the potential and complete possibilities of the web which improves it and provides interoperability" (Smith, 2004).

¹ Associate Professor of Information Science and Librarianship, Shiraz University, Iran

² M. A. In Information Science and Librarianship

Although the survey of the links has mostly been done in academic and scientific environments and the present study is not an exception, techniques of webometrics can be used in other atmospheres of the web such as, newspaper websites (Koosha, 2004), commercial websites (Voughan & Gao, 2005), newsagents' websites (Asnafi, 2006). Celvall & Haris (2003) have introduced three kinds of these researches for the study of webometrics and links:

- 1- Classification: Kinds of pages that have created links.
- 2- Stimulation: Reasons why individuals connect to these pages.
- 3- Relational: Relation between the researches done in universities and others' inclination in connecting links.

The present study is a relational one.

Influential Factors in Connecting Links:

Link rates are calculated by search engines. With the increase in the rate of search engine indexes this rate also increases. However, with passage of a longer time since the launching of a website the rate of links to that website will also increase. Links which have recently been created, that is, created in a shorter time than required for search engines to index the web, will not enter search engine indexes. Another factor in connecting links is URL. When the URL of a website changes, visibility of the website decreases which in turn reduces its link rates.

Differences and Similarities of Link and Citation:

The outset of the link theory is citation and citation analysis field. There are differences and similarities between link and citation. Links given to websites might be with reasons other than citation to scientific content, while citation is only with scientific reasons (Noruzi, 2005). For instance, there might be links from universities or journals' directories to other universities or journals which is not citation. Shaw points out that meta-links can be bi-dimensional, that is, web pages can connect each other without regarding their launching date, while citations are one-dimensional, that is, from the one who cites to the one being cited and not vice versa. However, because of invisible colleges he believes that a citation to others is possible in writers articles (Li, 2003). In order to understand stimulations behind connecting links we must pay attention to studies of link analysis. In his studies Kim mentions a number of stimulations in connecting links, which follow: publicity, validity of organization or institute, providing quick access to websites, providing graphic images, etc. On the other hand, Celvall divides the stimulations behind connecting links into four categories: ownership, social, general navigation, and gratuitous. Smith (2004) mentions 20% of links as similar to citation in printed articles. He clarifies the reasons behind connecting links by dividing web pages into different kinds. He believes that we should consider features of source pages (pages which have created links), features of intended pages (pages which have had links to) and reasons behind connecting links. Chu has analyzed links of university websites and concludes that 27% of the links had scientific stimulation and 50% of them were from information sources or directories.

In their article Vaughan and Shaw (2003) have concluded that in 57% of journals web citations have a meaningful correlation with catalogued bibliographic citations in citation index of Social Sciences and journal impact factor in ISI. They also believe that web citations are comparable and even replaceable by bibliographic citations.

Impact factor is different in the web and journals. There is no time period in web impact factor; on the other hand, a snapshot of the web at a certain time which has been achieved by search engines is important. In comparison with the content of a journal article, the content of a web source (not considering valid ones) has no possibility of judgment and survey, thus, it does not have quality control.

Limitations and Problems of Research in Webometrics:

Although commercial search engines are free, most of them do not allow access to about 200-400 of the results to their users. Search engines have internal problems. For example, their coverage is limited. In addition, invisible web is something search engines do not have. Vagueness of their search algorithm is another concern of search engines. Thus, the analytic results of the web performed by these search engines is only a glimpse of the results not an absolute one (Li, 2003). As a result, Bar-Ilan encourages web researchers towards self design crawlers. In addition, the depth of the indexing of search engines is important and influential in deciding the number of pages of a website. SocSciBot, working under the supervision of virtual statistical research group managed by Celvall, performs a free statistical study for all those who have sources for the web crawlers (Bar-Ilan, 2005). This software is in the field of Social Sciences and is a crawler of link calculations and has been created in 2001 with non-commercial scientific ends and with the aim of overcoming the search engines' problems. This crawler starts from the home page of the university website and then calculates all the links and internal pages of the same website.

An issue which has influenced webometrics research and issued criticisms towards it is that existing search engines do not cover the whole web. According to Antonio Gulli's studies most commercial search engines, as a whole, index 85% of the invisible web. Commercial search engines are limited in their coverage and are used in search ranking and not created for studies in webometrics (Noruzi, 2005). If we also consider the invisible web a great sum will be added to this amount.

Significance of the Problem:

Egghe considers web meta-links as very different from citation in scientific articles and Van Rann regards meta-links as superficial. In websites, links are mostly used for more observation of websites and also a better coverage of search engines, and thus, a better ranking of the search results. For webmasters and web designers surveying the connected links of a website is very significant. Through surveying the links they identify their audiences throughout the world, and thus, increase the visibility of websites. On the other hand, according to performed researches there is a meaningful correlation between the rate of research done at a university and the rate of connected links in those researches. Although in the performed studies no scientific relation has been found between connected links and research the significance of the problem will not decrease.

Role of University Rankings:

Besides informing the society of the function of universities, rankings of universities serves as a means of comparing universities, and thus, creating a competition between them. Following are a number of reasons for university rankings:

- 1- Informing the society of the status of academic institutes.
- 2- Improving a positive competition between academic institutes.
- 3- Encouraging academic institutes to continue development.
- 4- Pointing at the significance of knowledge in budgeting.
- 5- Attracting individuals' attention for continuing their education at chosen universities.

International Ranking Expert Group has always tried to improve the function of rankings by providing their viewpoints. This group by creating a logical structure in accepting current approaches and their league table, and participating in the analysis of tables and ranking approaches have improved present charts, and thus, enriched future rankings. Subjective nature of weighing, new international rankings, and evaluations which satisfy the needs of a larger category have been of great importance to this expert group. However, the used criteria, their weighing, and validity of the current indicators in ranking have made it possible to criticize or doubt university rankings.

Rankings performed at the three levels of local, national, and international have raised a large dispute. University rankings of the US, England, Poland, Germany, China, Australia, New Zealand, and Spain at a national level are some cases of rankings. These rankings have shown many weaknesses and strengths. Probably, the weaknesses of the internal methodology of rankings have underrated the strengths. Although these rankings do not reflect the scientific status of the universities, they influence their functioning by starting a competition between them regarding the criteria and indicators of ranking. Thus, the two rankings chosen are Academic Ranking of World Universities (ARWU) presented by Shanghai University and Times Higher Education Supplement (THES), which are the most well-known international rankings.

Shanghai University Ranking:

This ranking has been prepared by Shanghai Jiao Tang University. The aim of this ranking is to clarify the gap and difference between China's universities and the top universities of the world. Since this ranking has been prepared based on academic interests and does not have any foreign sponsor it is without bias in choosing the world top universities. About 500 top universities have been chosen.

Shanghai University does not consider this ranking as objective and announces some of its methodological problems as follows: quantitative evaluation vs. qualitative evaluation; research evaluation vs. education evaluation; variety of institutes; and language bias in publication. On the other hand, among the 20 universities having the first rankings 17 are from

the US. In calculating the total grade of published articles in three indexes a higher importance was given to documentation index of Arts and Social Sciences.

Shanghai's ranking indicators are:

- 1- Educational Quality (Alumni Groups—members of the institute who have won the Nobel Awards) = 10%
- 2- Faculty Quality: a) members of the institute who have won the Nobel Awards = 20%; b) HiCi—researchers who have been quoted to in 21 main topics = 20%
- 3- Research Products: a) published articles in Nature & Science = 20%; b) indexed articles in Social Sciences documentation index = 20%
- 4- Institute Size (functioning of the institute considering its size) =20%

Overall = 100%

This ranking is international and unified. Combining and separating its criteria does not change the results. It is a numerical ranking published annually and its sources can be found in the reports published by Shanghai University.

Times Ranking:

This ranking is also international and unified and combining or separating its criteria does not change its results. It is a numerical ranking published annually since 2004.

Times' ranking indicators are :

- 1- Peer review Score = %40
 - 2- Recruiter review = %10
 - 3- International faculty score = %5
 - 4- International student score = %20
 - 5- Faculty/Student score = %20
 - 6- Citations/Faculty score = %20
- Overall score = %100\

David Jobbins notes that the problem of this ranking is being business-based and priority for customers use. It seems that there is tangible overlap between these two rankings and the results of this research indicates that there is also a correlation between these two rankings.

Literature Review:

Thomas & Willet (2000, quoted in Koosha, 2003) compared counts of links of the web with ranking of educational departments in England. Thelwall (2000a) found out that Web Impact Factor (WIF) had correlation with university ranking. In this research, he found that when universities compare instead of educational departments, research tended to get a better result. Thelwall in England (2001a), Smith & Thelwall in Australia (2002) Tang and Thelwall in China (2002) Thelwall & Tang (2002), Koosha & Horri in Iran (2005) and Vaughan & Thelwall in

Canada (2005) found similar results. Other studies show that there was even significant correlation between external links and scholarly communications and research products in academic spheres in the web (Thelwall 2001a, Thelwall & Harris, 2003). Thelwall (2001) found significant correlation between link connect in 25 British universities and research product average in those universities. That was the first evidence showed there's relationship between link count to universities and their rankings. Smith & Thelwall found second evidence in academic sphere in Australian universities. In this study, number of members in scientific academic board stated as university size which is replaced with university host pages in Denominator which was used for WIF by Ingwersen (1997).

Research questions:

This Study tries to answer to these questions:

- 1- Is there any relationship between world university ranking and count of links to them?
- 2- Has connecting links to university websites the most correlation with indicators?
- 3- How much is the correlation between these two rankings? Is there any overlap between these two rankings?

Research Methodology:

a. Population and Sample: The population of Shanghai University ranking is 500 best universities and the size of the sample which is selected with Cochran's formula is 217 top universities. The size of sample is shown below :

$$\frac{\frac{t^2 * p * q}{d^2}}{1 + \frac{1}{N} (\frac{t^2 * p * q}{d^2} - 1)} = \frac{\frac{1.960^2 * .5 * .5}{.05^2}}{1 + \frac{1}{500} (\frac{1.960^2 * .5 * .5}{.05^2} - 1)} = 217$$

Times Higher Education (THE) ranking is used to measure accommodation and correlation with Shanghai ranking. Thus, for accommodation measurement, 100 top universities in Shanghai ranking were selected, then correlation between overall scores for universities in Times and Shanghai ranking was calculated:

b. Strategies for link connection : For the evaluation of external links established to the universities' website the below search strategy was used:

(Linkdomain:www.xxx.ext OR linkdomain:xxx.ext) AND NOT (host:www.xxx.ext OR host:xxx.ext)

For the assessment of internal links established to universities websites the below search strategy was used:

(linkdomain:www.xxx.ext OR linkdomain:xxx.ext) AND (host:www.xxx.ext OR host:xxx.ext)

Finally, for the extraction of host pages which a website possess the below search strategy was used:

host:www.xxx.ext OR host:xxx.ext

Data assessment was accomplished in May 2006. Altavista commercial search engine, text version, is used for the research shown in the below URL:

Altavista: Address [http://www.altavista.com/web/text]

This search engine was used because it could search all languages and was available and usable. Also, other researchers often used this search engine. Mirror sites and URL changes in university websites were not considered. In this research, SPSS 11.5 was used for statistical test. Pearson correlation test was used for the 1st and the 3rd questions and regression was used for the 2nd question.

Research Findings:

Research findings is shown in two tables in the correlation matrix between ranking indicators and external and internal links, host pages and correlation between two rankings.

Table1. Correlation matrix between ranking indicators and external & internal links and host pages

Variable	1	2	3	4	5	6	7	8	9	10
1	1	-	-	-	-	-	-	-	-	-
2	/712** 0	1	-	-	-	-	-	-	-	-
3	/500** 0	/781** 0	1	-	-	-	-	-	-	-
4	/554** 0	/836** 0	/735** 0	1	-	-	-	-	-	-
5	/729** 0	/876** 0	/535** 0	/598** 0	1	-	-	-	-	-
6	/670** 0	/913** 0	/629** 0	/676** 0	/865** 0	1	-	-	-	-
7	/533** 0	/733** 0	/467** 0	/372** 0	/674** 0	/662** 0	1	-	-	-
8	/518** 0	/855** 0	/660** 0	/740** 0	/722** 0	/797** 0	/534** 0	1	-	-
9	0/159*	050/0-	059/0	015/0	061/0-	042/0-	091/0-	101/0-	1	-
10	/588**	/478**	/355**	/346**	/498**	/437**	/368**	/354**	097/	1

0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---

** Correlation is Significant at the level 0/01(2-tailed).

* Correlation is Significant at the level 0/05 (2-tailed).

1=External links; 2=Overall score in Shanghai ranking; 3=Alumni winning Noble prize; 4=Staff winning Noble prizes and Field medals; 5=Highly Cited researchers in 21 broad subject categories; 6=Articles published in Nature and Science; 7; Articles indexed in SCI; 8=Size of institute; 9=Self links; 10=Host pages.

Table2. Correlation between Shanghai and Times rankings

Times ranking	Shanghai ranking	
0/842**	1	Shanghai ranking
1	0/842**	Times ranking

**Correlation is Significant at the level 0/01(2-tailed).

* Correlation is Significant at the level 0/05 (2-tailed).

Findings show that correlation between Shanghai ranking and external link connection established is 0.712 and correlation between this ranking with university host pages is 0.588 (both at the level 0.01) but correlation between WIF and Shanghai university ranking was not significant (Table 1). For the second question, research found that among Shanghai ranking indicators, the highest correlation was between external links (which is the most important links) and highly cited researchers in 21 broad subject categories (0.729), articles published in Nature and Science (0.670), staff winning Noble prizes and field medals (0.554), articles indexed in SCIE and SSCI (0.533), academic performance with respect of size of an institution (0.518), alumni winning Noble prizes and field medals (0.506). Findings also show that highly cited researchers in 21 broad subject categories has the most correlation with external links to university websites. So, we can state that with respect of derivation of link theory from citation theory, correlation between links and citation is mostly an indicator of university websites. This question is examined with the regression test between the variables in the research. In the regression test in trying to find out the most effective independent variables among above mentioned variables, researchers concluded that, highly cited researchers in 21 broad subject categories variable with distinction indicator 0.531 among other variables has the most effect in connecting links of universities' website which confirms the correlation between this variable and links to the university website.

Reference:

Academic Ranking of World Universities—Methodologies and problems . Online : [<http://ed.sjtu.edu.cn/rank/file/ARWU-M&P.pdf>]

Asnafi, Amir-reza. (2006). "A study of collaboration among Iranian News Agencies Websites using webometrics methods." [Unpublished].

Bar-Ilan, Judit.(2005). "Expectations versus reality–Search engine features needed for Web research at mid 2005." *Cybermetrics*. Vol.9(2005).
Online:[<http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>]

Bjorneborn, Iennart (1997). "Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach." Chapter2: Webometrics, p.14.
Online:[www.db.dk/lb/phd/phd-thesis.pdf]

Koosha, Keyvan (2003). "A Comparison of Iranian newspaper websites with the application of Web Impact Factor (WIF)." *Ettela'Shenasi*. 1(2:):87-114.

Koosha, Keyvan (2005). "Webometrics and Scholarly Communication: An Overview." *Faslnameh Ketab*. Quarterly Journal of the National Library of Iran, 14(4):7-16. Online:[<http://www.webology.ir/PLISA/articles/Koosha3.html>]

Koosha, Keyvan & Abbas Horri (2005). "The relationship between publishing and the count of academic In links to Iranian university web sites : exploring academic link creation motivations." Online: [<http://www.koosha.tripod.com/articles/issaward.pdf>]

Koosha, Keyvan & Mike Thelwall. (2005). "Motivation for URL citations to open access library and information science articles." 10th Conference of the International Society of Scientometrics and Information (ISSI). 24-28 July 2005, Stockholm, Sweden.

Li , Xumei . (2003). "A review of development and application of the web impact factor."Online *Information Review*. V.27 , N.6 pp. 407-417. Online:[DOI: 10.1108/14684520310510046]

Merisotis , Jamie & Jan Sadlak . (2005). "Higher education rankings evolution, Acceptance and Dialogue." *Higher education in Europe*. V.30 N.2.

Noruzi , Alireza . (2005). "Web Impact Factor for Iranian Universities." *Webology*. V.2 , N.1 . Online [<http://www.webology.ir/2005/v2n1/a11.html>]

Smith , Alastair G. (2004). "Web link as analogues of citations." *Information Research*. V.9 , N.4 , July 2004. Online :[<http://informationr.net/ir/9-4/paper188.html>]

Smith, A.G., Thelwall, M. (2002). "Web Impact Factors for Australian Universities." *Scientometrics*. Vol. 54 No.1/2, pp.363-80.

Tang, R., Thelwall, M. (2002). "Exploring the pattern of links between Chinese university Web sites", Proceedings of the 65th Annual Meeting of *Journal of American Society for Information Science and Technology*. Vol. 39 pp.417-24

Thelwall , Mike . (2001). "Extracting macroscopic information from Web links." *Journal of the American Society for Information Science and Technology*.V. 52 Iss.13, pp 1157-1168.

Thelwall , Mike . (2003). "What is this link doing here ? Beginning a fine-grained process of identifying reasons for academic hyperlink creation." *Information research*. V.8 , N.3 , April 2003 .Online : [<http://informationr.net/ir/8-3/paper151.html>]

Thelwall , Mike . (2002) . “The connection between the research of a university and counts of links to its web pages : An investigation based upon a classification of the relationship of pages to the research of the host university.” *Journal of American Society for Information Science and Technology*. Vol.54,2003,S.594-602. Online:[http://www.scit.wlv.ac.uk/~cm1993/papers/2003_connection_between_research_and_links_preprint.pdf]

Thomas, O. & Willet, P. (2000). “Webometric analysis of departments of Librarianship and information science.” *Journal of Information Science*. 26(6):421-428.

University Ranking in 2005 Jiao Tong Shanghai ,China.
Online:[<http://ed.sjtu.edu.cn/ranking.htm>]

Vaughan , Liwen & Debora Shaw. (2003) . “Bibliographic and web Citations : What is the Difference?” *Journal of American Society for Information Science and Technology*. N.54 (14) :1313-1322.Online: [<http://www.interscience.wiley.com>].[DOI :10.1002/asi.10338.]

Vaughan , Liwen , Yijun Gao & Margaret Kipp . (2005) . “Why are hyperlinks to Business Websites created ? A content analysis.” Online:[http://www.cais-csi.ca/proceedings/2005/vaughan_2005.pdf]

Vaughan, L. & Mike Thelwall. (2005). “A modeling approach to uncover hyperlink patterns: the case of Canadian universities.” *Information Processing and Management: an International Journal*. v.41 n.2, p.347-359.