Annals of Library and Information Studies
Vol. 57, June 2010, pp.109-121

Web impact factor and link analysis of selected Indian universities

Samir Kumar Jalal[1], Subal Chandra Biswas[2] and Parthasarathi Mukhopadhyay[3]

[1]Assistant Librarian, Central Library, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, Email: jalal_gtz@yahoo.co.in

[2]Professor, Department of Library & Information Science, University of Burdwan, West Bengal, Email: scbiswas_56@yahoo.co.in

[3]Sr. Lecturer, Department of Library & Information Science, University of Burdwan, West Bengal, Email: psmukhopadhyay@gmail.com

Investigates the effectiveness and relevance of web impact factors (WIFs) for Indian universities' websites. Reviews web impact factor as to how this link-based metrics is developed and is applied. Reports a case study on universities in West Bengal. SocSciBot 3.0 is used to generate link data in order to develop/form micro-link topology under study. Result shows that all the NITs are closely related in the topology framework/their activities whereas nodes are not linked significantly for the case of state universities and central universities.

Introduction

The Web is a collection of webpages connected to each other using hyperlinks. By clicking on the hyperlinks, users can move from one webpage to another located on different servers. Research on link analysis dates back to 1995-96 by many researchers in the field of computer science[1], information science[2], and mathematics[3]. Rousseau's[4] information analysis on the web and Almind and Ingwersen's[5] concept of web impact factor (WIF) and Rodriguez's[6] web citation analysis are important studies in the area of web link analysis. Major concepts like various WIFs, i.e., WIF_{simple} defined as ratio of total links and total webpages, $WIF_{external,}$ i.,e., ratio of total external links and total webpages, $WIF_{self\text{-}links,}$ i.,e., ratio of total self-links and total webpages have been developed.

The relationship among universities can be measured and visualized through multiple indicators such as scientific collaboration, consortium approach, resource sharing etc. But, from webometric point of view, these relationships may be measured through hyperlink flows that points from one university to another. Studying the link relationship among universities in their academic web space would be quite interesting.

Web impact factors (WIFs), one of the webometric indicators aims to evaluate impact factor of specific websites or any top-level domains (TLDs), one of the domains at the highest level (e.g. .in) in the hierarchical Domain Name System (DNS) of the Internet or second-level domains (SLDs), domain (e.g. ac.in) that is directly below a top-level domain (TLD). It is relatively a new measure to the extent to which a website is linked to by other sites. Therefore, WIF is analogous to citation counts in print environment[7]. One obvious difference may be the nature of documents that are linked. Citations in conventional print publications are generally between research publications, while web links may be between wide varieties of publication types: personal home page, subject resource guide, research article, advertisement, etc. Smith[8] investigated the extent to which web links are analogous to the citations in traditional print literature and found that overall, 20% of the Web links in the study could be regarded as research links analogous to citations. However, anyone who can create a webpage and host it on the web could link to any other pages without quality control to that of scholarly journal[9] and in practice web links are not entirely equivalent to citations in the scholarly literature[10].

The Web impact factor (WIF) was developed by Ingwersen[11] to measure the impact of websites by the number of links it receives. He added that WIF can be defined as the ratio of links made to a website to the number of pages at the website. According to him, there are three types of WIF such as simple, overall and external WIFs. Among them, external WIF appears to be the best valid measures of impact factor. This is very much similar to Google's concept of page rank[12]. It is the extended concept of impact factor, which has been introduced by Garfield[13] and he also pointed out that the WIF is

analogous to the Citation Impact Factor (CIF) proposed by Eugene Garfield.

The present study is intended to analyze select Indian universities websites using web impact factors. Besides, various link analyses have been undertaken in order to know the pattern of linkages among the similar group of universities, e.g., among central universities, NITs, IITs, IIMs, state universities, etc.

The impact factor is a measure of frequency with which the average article in a journal had been cited in a particular year[14]. In the web environment, impact factor is measured through the number of hyperlinks counts and number of webpages. The concept of self-citation is replaced by self-links, i.e., the links within the websites and citation is replaced by in-links, i.e., the links coming outside the websites. As we know, WIF is the logical sum of external and self-link webpages divided by number of web pages found on that particular websites. Egghe[15] pointed out that citations are very different from hyperlinks. They may be synchronic while citations are diachronic. Only, previously published papers can be cited, not vice versa.

Literature review

The early development of the calculation of WIF did not yield satisfactory results[16-19]. The WIF generally is defined as the ratio between the number of total links and total webpages of a specific domain. In this case, the total number of links consists of internal links plus external links. For a large website having large number of webpages, there are a good number of internal links which are actually used to organize the internal pages and do not deserve any credit for the reputation of a website. Links from external pages gives more useful information and thus enhance the reputation of the institute. Thelwall[20] tried to extend the concept of impact factor for web-based resources and to use the power of search engines to cover other domains on the Internet. A survey was conducted in order to test the coverage of search engines to calculate WIF, which was found sometimes extremely uneven and leads to misleading calculation.

Smith and Thelwall[21] calculated Web Impact Factors (WIFs) for Australian universities using a specially designed crawler and the Altavista search engine. Links between UK, Australian and New Zealand universities had been reflected. Both the number of pages at the site, and the number of academic staff members, were used as measures of the size of the universities. The WIFs were compared with conventional measures of research output: rankings by Asiaweek magazine, the number of publications per staff member, and the number of citations per staff member. There is a good correlation between the crawler and Altavista in estimating the link counts. The WIFs do not appear to correlate well with conventional measures of research output. They also discussed some of the methodological issues in the calculation of WIFs.

Terveen and Hill[22] examined the number of hyperlinks between websites as an indicator of the quality of sites and found that hyperlink connectivity had a significant relationship to experts' quality judgment of sites. But, the statement, "WIF is a measure of the reputation of a website" may not always hold true due to WIF selflink. Thelwall[23] did a comparison of sources of links for academic web impact factor calculations. It has been demonstrated that several versions of the metric can produce results that correlate with research rating of British universities.

Thelwall[24] used academic Web crawler to study links to six UK universities. Smith and Thelwall[25] studied the links between UK, Australia and New Zealand universities where both crawler and Altavista were used. The number of academic staff members were used to represent the size of university replacing webpage as WIF denominator. But, Smith studied Australian universities and online journals using external inlinks as WIF numerator

Jalal, Biswas and Mukhopadhyay[26] had shown in their hyperlink study for the state universities of West Bengal that IIT Kharagpur occupied the first rank among the universities based on WISER indicators and Uttar Banga Krishi Vishwavidyalaya got the last position from the point of view of webometric ranking.

Mukhopadhyay[27] tried to explore the possibility of research in the field of webometrics in the educational institutions in India using Web Impact Factor (WIF).

Li[28] studied hyperlinks extensively by applying existing bibliographic methods and made an exhaustive review the development of WIF. Li pointed out the origin of WIF

and techniques for data collection using commercial search engines. The study also highlights the development of WIF- origin, traditional measures and its improvements.

Mukhopadhyay[29] reported the result of webometrics investigation at different levels of domain system. He had shown the calculation of web impact factor for Country Code for Top Level Domain (ccTLD) of South Asian Association for Regional Cooperation (SAARC) group of countries, second-level domain (SLD) related to education and research. He had also shown the ranking of Indian Institutes of Management (IIMs) and Indian Institutes of Technology (IITs) in India. Besides, Mukhopadhyay mentioned the problems and prospects of webometrics, as a new discipline.

Noruzi[30], in a study investigated the web impact factors (WIF) for Iranian universities and introduced a new system of measurement. WIF for Iranian universities were calculated by dividing link counts by the number of pages found in Altavista for each university at a given point of time. These WIFs were then compared to study the impact factor, visibility and influence of the Iranian universities' web sites. Over all Iranian universities' web sites have a low WIF while specific feature of sites may affect an institution's WIF.

Noruzi[31] investigated that the web presence and WIF for Country Code Top Level Domain (ccTLDs) of Middle Eastern countries based on Yahoo search. The WIF was calculated at to levels: Top Level Domains and Sub-Level Domains (SLDs). The results show that the Middle Eastern countries, apart from Turkey, Israel and Iran have a low web presence. On the other hand, their web sites have a low inlink WIF. It was pointed out that specific features of sites might affect countries WIF.

Amipour and Payam[32] studied the impact factor of 40 universities covered by Iranian Ministry of Health on the web. Accordingly, total number of web pages and links into each university were counted by advance search feature of Altavista search engine. The impact factors for overall, in-link and self-link for each web sites were calculated. Finally, the universities were ranked according to those parameters. The results show that although Teheran University of Medical Science is first in the number of web pages, the overall results suggest that Iranian universities of Medical Sciences are not well known internally due to their poor scholarly publications and focus on regional language in their web pages.

Elgohary[33] made a study in order to investigate the WIF of Arab universities. The study included 99 universities representing 20 Arab countries. The advanced search facility of AltaVista was used for data collection. Two rounds of data collection were conducted to retrieve the links as well as the web presence of the included universities. The findings revealed that Jordanian universities represent 40 percent of the top ten universities with the revised web impact factor. However, this was not the case in terms of the universities' web presence. Results indicated a strong correlation between external links and web presence.

Jayshankar and Babu[34] in a webometric study examined the websites of 45 universities in Tamil Nadu to analyse the number of webpages, links, calculate various types of WIFs. The result found that although some universities of Tamil Nadu have quite large number of webpages but very low number of inlinks and hence low WIF. Ravikumar[35] investigated the link pattern of selected academic libraries in India using UCINET computer software to visualize the network pattern that existed among peer group libraries.

In the recent years, clear evidence has emerged to show that counts of links to scholarly web spaces (universities and departments) can correlate significantly with research measures, giving some credence to their use for the investigation of scholarly communication[36]. The specific questions addressed here are whether site age and site content are inducers of links to a journal's web site as measured by the ratio of link counts to journal impact factors. The results show that both site age and site content are significant factors for the disciplines studied: library and information science, and law. Mukherjee[37] made a study to know scholarly impact of websites of central government universities using Google and MSN.

Objectives of the study

The primary objective of the study is to examine critically the effectiveness and efficiency of the use of web impact factor and to find out the link patterns among the selected universities under study. The other inter-linked objectives are as follows:

- To calculate various types of web impact factors for Indian universities;
- To investigate relevance of web impact factor (WIF) for judging web performance of the Indian universities;
- To analyze suitability of using various types of WIFs for ranking the Indian universities;
- To find out link patterns among the central universities, state universities and NITs in India; and
- To generate micro-link topology among similar kind of universities using appropriate webometric tools.

Methodology

Data collection

There are total 150 deemed universities (comprising of 130 universities as listed in UGC website, i.e., www.ugc.ac.in and 20 NITs, 255 state universities (list of state universities is mentioned in the UGC website), 23 central universities (excluding 16 new central universities established recently), 19 institutes of national importance, 13 open universities and 59 private universities in India as of March 2010. Data have been collected using AltaVista and Google search engines during the specified period as mentioned in the concerned tables.

It is clear from Table1 that only 17 universities out of 526 (i.e., 3.24%) universities do not own websites or the websites may be non-functional as of March 2010.

Selection of webometric tools

Webometric tools are the most important elements in webometric research. Webometric tools are generally of two types, i.e., commercial search engines and personal web crawlers. Most popular commercial search engines are Yahoo!, MSN, AltaVista, Google, AllTheWeb, etc. These commercial search tools are powered by special query syntaxes, which are explained below. Besides, these search engines also support Boolean, relational and positional operators (AND, OR, NOT, NEAR, ADJ etc.) for retrieving data. Webometric tools employed for the present study are both commercial search engines as well as personal web crawlers like SocScibot[38] for all the 23 central universities, 22 universities in West Bengal and 20 NITs in India. SocSciBot 3.0 has been used in this study because it is widely used and tested by many webometricians for their studies in academic institutions. Both the tools have been used to validate the results and also to find out a microstructure of web relationships within the selected sites. Another tool, used for visualization of links relationship, is Pajek[39]. This powerful visualization tool is embedded with SocSciBot. The information flow for visualizing the network diagram can be explained as in Figure1.

Table 1 — Various types of Indian universities having websites in March 2010

Type of universities	Total	Universities having websites		Sample chosen
		Total	Percentage	
Central university	23	23	100	23 (100%)
Deemed university	150	148	98.66	20 NITs (100%)
State university	255	245	96.07	22 universities in W.B (100%)
Institute of national importance	19	19	100	Not selected
State legislative	7	7	100	Not selected
Open university	13	13	100	Not selected
Private University	59	54	92	Not selected
Total	526	509	96.76	

Source: www.ugc.ac.in & www.dde.ac.in

Fig. 1— Information flow for building the network diagram

Table 2 — Webometric query syntax with results

Search command	Results	Supported by
domain:du.ac.in	Total number of web pages for the University of Delhi	Google, AltaVista, Yahoo!
site:du.ac.in	Total number of webpages for the University of Delhi	Google, AltaVista, Yahoo!
linkdomain: du.ac.in –domain: du.ac.in	Total number of inlinks for the University of Delhi	AltaVista, Yahoo!
linkdomain:in domain:in	Total number of self-links for India	AltaVista, Yahoo!
linkdomain	Total number of links for India	AltaVista, Yahoo!
site:du.ac.in file:pdf	Reported total number of pdf files for University of Delhi	Google, AltaVista, Yahoo!

Use of appropriate query syntaxes

The Webometric analysis is based on the data collected from the Web using various search engines. In each search engine, there are some specific search keywords (called special keywords) that are designed to retrieve information from Web. These specific search keywords along with search syntaxes are given in Table 2 for ccTLD *.in* (for India) and the domain name du.ac.in (for University of Delhi).

Choice of indicators for ranking

The first web indicator, WIF was proposed by Almind and Ingwersen[40] based on link analysis that combines the number of external in-links and the number of pages of the website, a ratio of 1:1 between visibility and size. This ratio was modified later on by adding two new indicators to the size component, i.e., number of documents in rich file formats (formats that are generally in use for scholarly communications) and number of publications being collected by Google Scholar database. This new measure is called WISER. As it has been already commented, these four indicators can be measured from the quantitative results provided by the main search engines. The WISER ranking methodology[41] may be presented here as follows:

- Size (S). Number of pages recovered from four engines: Google, Yahoo, Live Search+ and Exalead. For each engine, results are log-normalized to 1 for the highest value. Then for each domain, maximum and minimum results are excluded and every institution is assigned a rank according to the combined sum.
- Visibility (V). The total number of unique external links received (inlinks) by a site can be only confidently obtained from Yahoo Search, Live Search and Exalead. For each engine, results are log-normalised to 1 for the highest value and then combined to generate the rank.
- Rich Files (R). After evaluation of their relevance to academic and publication activities and considering the volume of the different file formats, the following were selected: Adobe Acrobat (*.pdf*), Adobe PostScript (*.ps*), Microsoft Word (*.doc*) and Microsoft Powerpoint (*.ppt*). These data were extracted using Google and merging the results for each filetype after log-normalizing in the same way as described before.
- Google Scholar (Sc). Google Scholar provides the number of papers and citations for each academic domain. These results from the Scholar database represent papers, reports and other academic items.

These four ranks are then combined according to a formula where each one has a different weight:

+ Live search has withdrawn all link search operators except for *linkfromdomain*

Table 3 — Calculation of WIF for India for TLD (.in domain)

Search engines	Actual Data				WIF		
	Webpages	Inlinks	Self-links	Total links	Simple	external	self-link
AltaVista	132000000	36100000	11500000	47500000	0.36	0.27	0.09
Google	372000000	349000000	479000	326000000	0.88	0.94	0.00
Yahoo	760000000	35700000	11400000	760000000	1.00	0.05	0.02

Note: The data for the table has been collected during March 28-29, 2009 using AltaVista, Google and Yahoo!

Table 4 — Academic domain WIF for India for SLD (.ac.in domain)

Search engines	Webpages	Actual Data			WIF		
		Inlinks	Self-link	Total links	Simple	External	Self-link
AltaVista	2310000	434000	1080000	1130000	0.49	0.19	0.47
Google	7570000	26000000	369000	8880	0.00	3.43	0.05
Yahoo	1319835	445000	1100000	233,955	0.18	0.34	0.83

Note: The data for the table has been collected during March 28-29, 2009 using AltaVista, Google and Yahoo!

Table 5 — Average WIF for Indian Universities in 2009

Type of universities	Total number	No. of Univ. having websites	WIF-overall	WIF-external	WIF-self-links
Deemed university (excluding NITs)	130	123	0.57	0.27	0.44
NITs	20	20	0.42	0.11	0.62
State universities	255	245	0.97	0.54	0.52
W.B	22	22	1.17	0.43	0.56
Central universities	23	23	1.16	0.56	0.51

Source: Yahoo!, dated March 28 to April 6, 2010

Webometrics Rank (WR) [42] = 4*Rank V +2*Rank S +1*Rank R+1*Rank Sc,
where V= Visibility, S= Size, R= Rich Files and Sc= Google Scholar

The volume of contents is measured by the number of pages freely accessible and their visibility by the number of incoming links. The number of rich files is used as an indicator because rich files are preferred formats for scholarly communications. Total number of documents indexed in Google Scholar is also considered as an important indicator for scientific publications on the Web. Each web domain is ranked by the linear aggregation of these indicators for their ranking.

Calculation of Web Impact Factors

WIF is the web versions of impact factor. There are three types of WIFs: WIF-simple, WIF-revived and WIF-overall. The impact factor is a measure of frequency with which average article in a journal had been cited in a particular year or period. WIF is the ratio of the number of backlinks to a site, divided by the number of webpages at the site.

Let,

A = Total number of webpages to a particular site

B = Number of external backlinks to a given site

C = Number of self-links to a given site

D = Total number of links to a site

Therefore, WIF_{simple}= D/A; $WIF_{Revised}$ =B/A and $WIF_{selflink}$ = C/A

Ranking of universities

Ranking of higher educational institutes is gaining importance in the context of globalization. Therefore, an efficient ranking system helps all the stakeholders, parents, teachers, administrators and funding agencies. In the age of digitization, much interest has grown to rank the academic institutions based on web visibility and performance of their websites. The webometrics approach to rank institutions follows the Berlin Principles for ranking institutions in which it is advised that appropriate weightages should be assigned to various indicators like size, visibility, Rich files and Google scholar. Webometrics group used the following formula to rank the institutions:

Webometrics Rank (position) = 4*Rank V +2*Rank S +1*RankR+1*Rank Sc, where V= visibility, S= Size, R= Rich Files and Sc= Google Scholar

Ingwersen[43] pointed out that while calculating WIF great care is to be taken for the institutional website because results may fluctuate, whereas the same may be more stable for TLDs and SLDs. Another observation is that value of WIF does not significantly correlate with research rating. Smith[44] studied Australian universities and online journals using external inlinks as WIF numerator. Here number of academic staff member was used to represent the size of university replacing webpage as WIF denominator. Jalal, Biswas & Mukhopadhyay[45] made a webometric study for the IITs and IIMs and results show that IIT Bombay occupied the first place based on WIFs as well as WISER indicators. Many webometricians thought that internal links have a great influence on the calculation of WIF but internal links are created with the objective of organizing internal webpages in a website. It has been observed from a dataset of more than 500 Indian universities that those universities, which are having more webpage, have less value of WIF.

The WIF gives a measure of average impact per page and this concept can be applied to either university or institutions, sub-level domain or top level domain, e.g., for countries (ccTLDs). The WIF measure is not free from bias because the same formula is applied to all types of websites ranging from 1 to more than 1,00,000. Secondly, in WIF calculation, the new and old websites are considered equal. In such a situation, the external links will be more in case of older websites than the younger one resulting into the huge change in the value of revised WIF or WIF-external. Another influencing factor to be noted is that if we want to calculate WIF for some universities or institutions having a wide variation of webpages and links then some kind of normalization is required first and then WIF formula should be applied to rank web sites. Webpage inlink rates determine the website impact factor but not vice versa. Mathematically, we can say that WIF = f(r), where r=> rate of external links i.e. B/A.

Alternative document model (ADM)

There is a need for collecting data not only through commercial search engines but also academic web crawlers to overcome any biases. The advantage of academic crawler is that it is possible to cover individual website comprehensively within specified parameters[46]. It is not possible for web crawlers to cover large web areas for which search engines are most appropriate. With much criticism for the data collected from search engines, Thelwall suggested Alternative Document Model (ADM) to collect the link data for universities.

ADM and link analysis for national institute of technology (NIT) of India

ADM Counts summary for NITs in India during Aug 2009 is been reflected in Table 6.

The ADMs have four parts generally:

- Page level: original link data is transformed into page link data by truncating the URLs;
- Directory level: original link data is transformed into directory link data by truncating before the last slash of the URLs of all pages;
- Domain level: link data is transformed into domain link data by truncating the page and link urls after the first slash following the domain name.
- Whole university model: whole university will be regarded as the unit to count the links

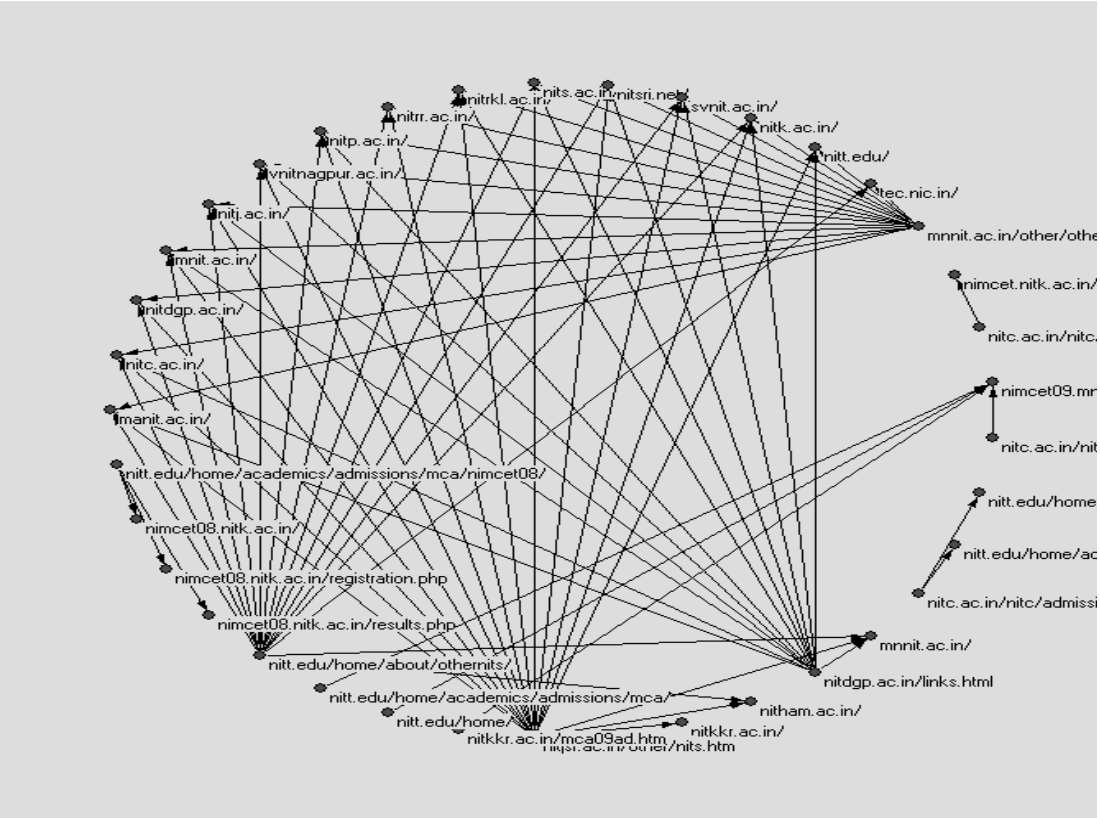

Fig 2 — Link topology for National Institute of Technologies in India in 2009

ADM and link analysis for central universities of India

The raw data is collected using personal web crawler, i.e., SocSciBot 3.0 for central universities and the ADM count summary for central university is explained in Table 7.

It has been found from Figure 2 that sites of central universities in India are not well connected. Few bilateral links have been observed from figure-3, for example there are hyperlinks between Visva-Bharati University in West Bengal and Tezpur University, Assam. Topologies of link relationships are also found between Delhi University and Jawaharlal Nehru University (JNU) and also Banaras Hindu University and University of Allahabad.

ADM and link topology for universities in West Bengal

The ADM counts for the state of West Bengal universities are mentioned in Table 8. The details in the table are extracted using SocSciBot 3.0

Figure 4 shows the micro-link topology formed among the universities in West Bengal.

It has been found from the above figure that there is inter-relationship among Indian Statistical Institute, Kolkata, National Institute of Technology, Durgapur and Indian Institute of Technology, Kharagpur and interestingly all these three link nodes are central institutes, not the state universities.

Findings and discussions

SosSciBot 3.0 has been used to extract the link data for the purpose of constructing link topology formed for NITs in India as shown in Fig 2. It clearly reflects that there is much closeness among the NITs under study. The reason may be due to similarities of the objectives of the institutes and the courses conducted. NIT Durgapur, NIT Allahabad, NIT Calicut, NIT Surathkal and NIT Trichi are producing maximum inlinks. On the other hand, Figure 3 reflects that there is no significant web-based associations amongst central universities in India but Delhi University, Banaras Hindu University, Jawaharlal Nehru University, Aligarh Muslim University are having maximum inlinks and connected with each other. Fig 4 shows that link topology among universities in West

Table 6 — ADM counts summary for NITs in India during Aug 2009

Name	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
vnitnagpur.ac.in	5	5	5	5	0	0	0	0
nitrr.ac.in	3	3	3	3	0	0	0	0
nitp.ac.in	4	4	4	4	0	0	0	0
nitsri.net	5	5	5	5	0	0	0	0
tec.nic.in	6	3	3	3	0	0	0	0
nitt.edu	6	6	5	5	21	19	18	16
nitj.ac.in	5	5	5	5	0	0	0	0
nitdgp.ac.in	4	4	4	4	13	13	13	13
mnnit.ac.in	8	8	7	6	15	15	15	15
nitjsr.ac.in	0	0	0	0	18	18	18	18
nitkkr.ac.in	1	1	1	1	1	1	1	1
manit.ac.in	4	4	4	4	0	0	0	0
nitc.ac.in	5	5	5	5	4	4	3	3
mnit.ac.in	5	5	5	5	0	0	0	0
svnit.ac.in	5	5	5	5	0	0	0	0
nitw.ac.in	1	1	1	1	0	0	0	0
nits.ac.in	5	5	5	5	4	1	1	1
nitk.ac.in	9	7	7	6	0	0	0	0
nitham.ac.in	2	2	2	2	0	0	0	0
nitrkl.ac.in	4	4	4	4	11	11	11	11

Source: SocSciBot 3.0 Aug 1-7, 2009

Table 7 — ADM count summary for central university

Name	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
www.amu.ac.in	0	0	0	0	0	0	0	0
allduniv.ac.in	1	1	1	1	0	0	0	0
jmi.nic.in	0	0	0	0	1	1	1	1
mzu.edu.in	0	0	0	0	0	0	0	0
ggsipu.nic.in	0	0	0	0	0	0	0	0
bbauindia.org	0	0	0	0	0	0	0	0
nehu.ac.in	0	0	0	0	0	0	0	0
du.ac.in	2	2	2	2	0	0	0	0
assamuniversity.nic.in	0	0	0	0	0	0	0	0
jnu.ac.in	1	1	1	1	1	1	1	1
pondiuni.edu.in	0	0	0	0	0	0	0	0
tripurauniversity.in	0	0	0	0	0	0	0	0
sikkimuniversity.in	0	0	0	0	0	0	0	0
rgu.ac.in	0	0	0	0	0	0	0	0
uohyd.ernet.in	0	0	0	0	0	0	0	0
visva-bharati.ac.in	1	1	1	1	0	0	0	0
manipuruniv.ac.in	0	0	0	0	0	0	0	0
hindivishwa.org	0	0	0	0	0	0	0	0
nagauniv.org.in	0	0	0	0	0	0	0	0
bhu.ac.in	0	0	0	0	2	2	2	2
manuu.ac.in	0	0	0	0	0	0	0	0
tezu.ernet.in	0	0	0	0	1	1	1	1
efluniversity.ac.in	0	0	0	0	0	0	0	0

Source: SocSciBot 3.0 dated May 17-18, 2009

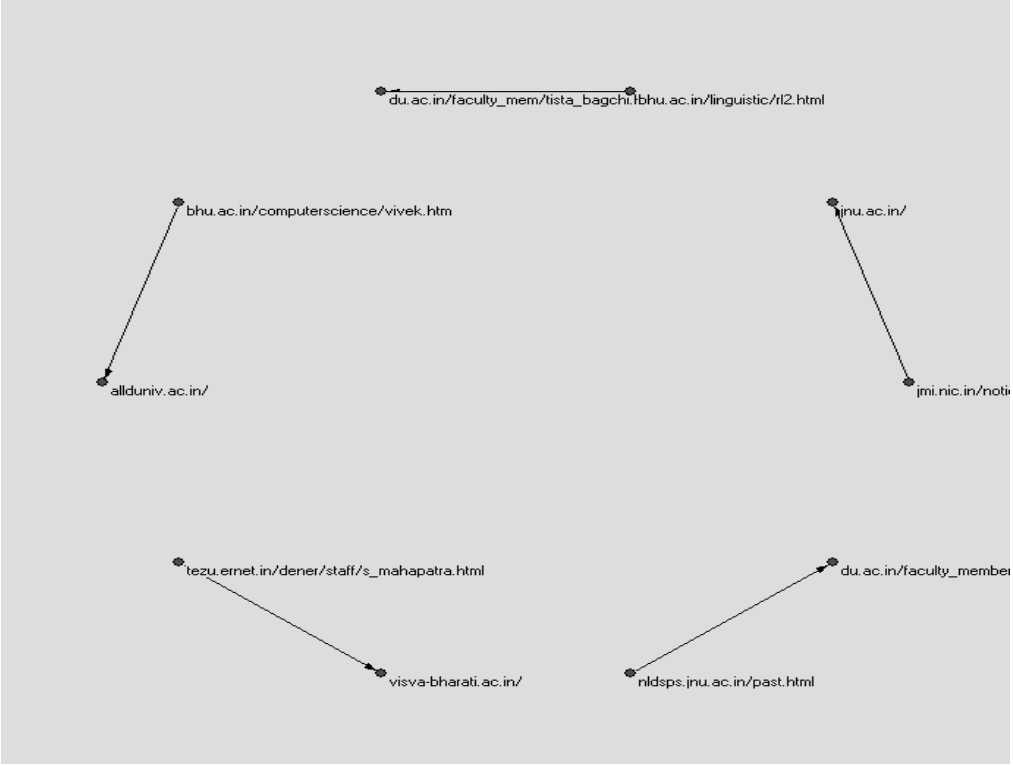

Fig 3 — Link analysis for central universities in India

Bengal revealed very poor results. It is found that state universities in West Bengal are not connected with each other in web space; of course, ISI Kolkata, NIT Durgapur and IIT Kharagpur formed a micro link topology showing the existence of their connection with respect to hyperlink analysis.

Webometric analysis is based on the web-based datasets. These datasets are derived from the web with the help of either commercial search engines or personal web crawler. The popular personal web crawlers are LexiURL[47] and SocSciBot. There are many challenges associated with the collection of data with respect to the dynamic nature of the Web. Bray[48] raised the question, *"How big is the Web? And what is the average page like?* The following query syntax will help to retrieve data from the web for the calculation of Web Impact Factor (WIF) such as Simple-WIF, Self-link-WIF and external link-WIF. Besides, it has been experimented that instead of using Boolean operation, '-'instead of NOT and'-', instead of AND gives more reliable results. For example in case of determining self-link data, syntax may be as follows: *linkdomain:isical.ac.in domain: isical.ac.in.* For extracting external links, the query may be: *linkdomain:isical.ac.in -domain:isical.ac.in.*

Now-a-days, university websites are increasingly used for a wide variety of purposes, from attracting new students to providing online library catalogues; from attracting research projects to communicating the research findings; from announcing the achievements of individuals, research groups, institutes and departments to hosting online articles or by publishing summaries, etc. In such a situation, there are lot of changes taking place over time with respect to counting the link data. Payne[49] made an exhaustive study of academic web links to identify quantum of change and its reasons. The changes

Table 8 — ADM count summary for universities in West Bengal

Sl No	Name	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
1	buruniv.ac.in	0	0	0	0	0	0	0	0
2	caluniv.ac.in	0	0	0	0	0	0	0	0
3	becs.ac.in	0	0	0	0	0	0	0	0
4	bckv.edu.in	0	0	0	0	0	0	0	0
5	iitkgp.ernet.in	1	1	1	1	0	0	0	0
6	isical.ac.in	1	1	1	1	0	0	0	0
7	nitdgp.ac.in	0	0	0	0	2	2	2	2
8	wbnsou.ac.in	0	0	0	0	0	0	0	0
9	rabindrabharatiuniversity.net	0	0	0	0	0	0	0	0
10	rkmvu.ac.in	0	0	0	0	0	0	0	0
11	klyuniv.ac.in	0	0	0	0	0	0	0	0
12	nbu.ac.in	0	0	0	0	0	0	0	0
13	ubkv.ac.in	0	0	0	0	0	0	0	0
14	vidyasagar.ac.in	0	0	0	0	0	0	0	0
15	visva-bharati.ac.in	0	0	0	0	0	0	0	0
16	nujs.edu	0	0	0	0	0	0	0	0
17	thewbuhs.org	0	0	0	0	0	0	0	0
18	wbut.net	0	0	0	0	0	0	0	0
19	jadavpur.edu	0	0	0	0	0	0	0	0
20	iimcal.ac.in	0	0	0	0	0	0	0	0
21	ugb.ac.in	0	0	0	0	0	0	0	0
22	aliah.ac.in	0	0	0	0	0	0	0	0

Source: SocSciBot 3.0 dated May 15-16, 2009

may be due to incorporating a new large collection of pages.

Conclusion

Hyperlink analysis is basically applied to web pages or websites to retrieve information and also collect data for the purpose of ranking. Different ranking methods used this approach as their strategies. It has been observed from the above study that the NITs are densely linked with each other whereas it is not the case of central universities or state universities of West Bengal. Therefore, it can be remarked that the nature of universities and their work phenomenon is closely related to number of external link counts. Inter-institutional collaboration research works or exchange programmes may have a strong impact on web relationship among the universities.

Finally, we may conclude that although WIF has been widely used as webometric indicators to judge the quality of websites through the value of its co-efficient, it does not in most cases reflect reliable results. WISER indicator comprising of components from WIFs and two other components, i.e., Rich Files and Google scholar with appropriate weights has been combined to rank the institutions in this study. However, further research is required to improve the ranking methodology by incorporating more variables like infrastructure, number of permanent faculty members, number of students, number of publication of the staff members in reputed foreign journals, student placement etc.

Acknowledgements

We are very grateful to anonymous reviewers for making helpful comments towards the improvement of the paper.

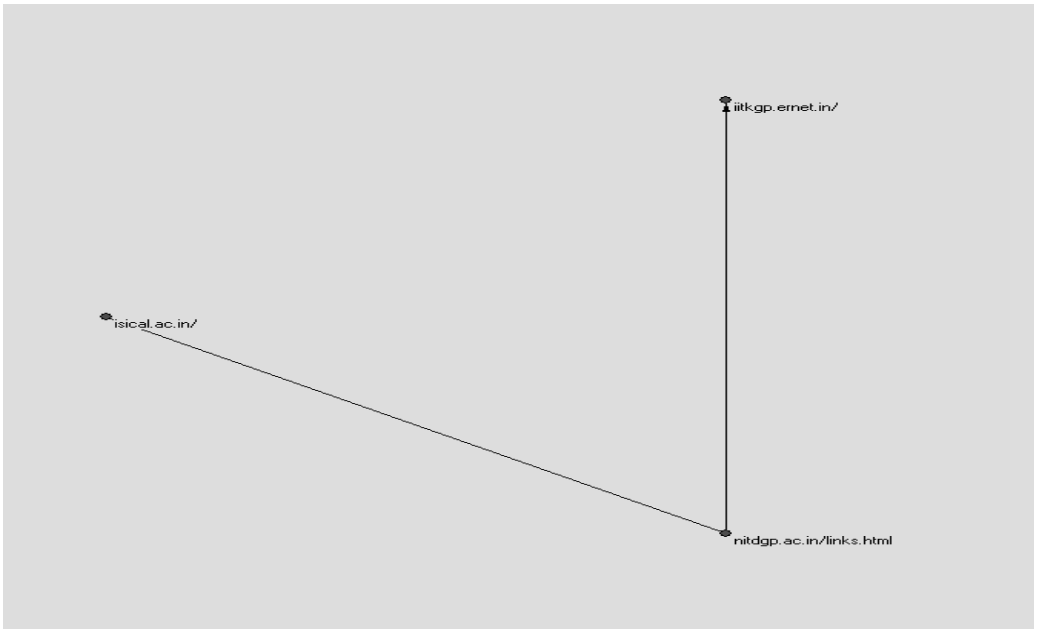

Fig 4 — Link analysis among universities in West Bengal

We are also thankful to the editors for their constructive suggestions in improving the paper.

References

1. Weiss R, Velez B, Sheldon M, Manprempre C, Szilagyi M, Duda A and Gifford D, HyPursuit: A hierarchical network search engine that exploits content-link hypertext clustering, *Proceedings of the 7th ACM conference on Hypertext,* ACM Press: New York, 1996
2. Bossy M J, The last of the litter: Netometrics. Available at: http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html (Accessed dated 16th March, 2010)
3. Abraham R H, Webometry: measuring the complexity of the World Wide Web. Available at: http://www.ralph-abraham.org/vita/redwood/vienna.html (Accessed dated 16th March, 2010)
4. Rousseau R, Sitation: an exploratory study, *Cybermetrics*, 1(1) (1997) paper-1. Available at: *http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html (Accessed dated 15th March, 2010)*
5. Almind T C and Ingwersen P, Informetric analyses on the World Wide Web: Methodological approaches to Webometrics, *Journal of Documentation,* 53(4) (1997) 404-26.
6. Rodríguez, i Gairín, JM, Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la red. [Impact assessment of information on the Internet: AltaVista, the Citation Index of the Web]. *Revista Española De Documentación Científica*, 20(2), 1997, 175-181. Retrieved May 10, 2010, from http://bd.ub.es/pub/rzgairin/altavis.htm Impact assessment of information on the Internet: AltaVista, the citation index of the Web. Available at:www.kronosdoc.com/publications/altavis.html.
7. Davenport E and Cronin, B, The citation network as a prototype for representing trust in virtual environments. In Cronin, B. (Eds), *The Web of Knowledge: a Festschrift in Honour of Eugene Garfield*, *Information Today*, Medford, (2000) 517-34.
8. Smith A G, Web links as analogous to citations, *Information Research,* 9(4), July 2004. Retrieved dated 13th Dec 2009 from http://informationr.net/ir/9-4/paper188.html
9. Vaughan L and Thelwall M, Scholarly use of the web: What are the key inducers of links to journal web sites*? Journal of the American Society for Information Science and Technology,* 54(1) (2003) 29-38.
10. Smith A and Thelwall M, Web impact factors and university research links. *Proceedings of the 8th International Conference on Scientometrics and Informetrics,* Sydney Australia, 16-20 July 2001, 2 (2001), p.657-64.
11. Ingwersen P, The calculation of Web Impact Factors. *Journal of Documentation*, 54(2) (1998) 236-43
12. Brin S and Page L, The anatomy of a large-scale hyper textual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7) (1998) 107-117.
13. Garfield E, Fortnightly Review: How can impact factors be improved? *British Medical Journal,* 313 (1996) 411-413.
14. Garfield E, The impact factor, *Current Contents*, (1994) June 20.
15. Egghe L, New informetric aspects of the Internet: some reflections and many problems, *Journal of Information Science*, 26(5) (2000) 329-335.
16. Smith A, A tale of two Web spaces: comparing sites using Web impact factors, *Journal of Documentation*, 55(5) (1999) 577-592.
17. Thelwall M, Conceptualizing documentation on the web: an evaluation of different heuristic-based models for counting links between university websites. *Journal of the American Society for Information Science and Technology*, 53(12) (2001) 995-1005.

18. Thomas O and Willett P, Webometric analysis of departments of librarianship and information science, Journal of Information Science, 26(6) (2000) 421-28
19. Björneborn L and Ingwersen P, Perspectives of Webometrics, *Scientometrics*, 50(1) (2001) 65-82.
20. Thelwall M, Web impact factor and search engine coverage, *Journal of Documentation,* 56(2)(2000) 185-189
21. Smith A and Thelwall M, Web Impact Factors for Australasian universities, *Scientometrics*, 54(1/2) (2002) 363-380.
22. Terveen L and Hill W, Evaluating emergent collaboration on the Web, *Conference of computer supported Cooperative work, Seattle, Washington,* November, 1998
23. Thelwall M, A comparison of sources of links for academic Web Impact Factor calculations, *Journal of Documentation*, 58 (1) (2002) 60-72.
24. Thelwall M, Results from a Web Impact Factor crawler, *Journal of Documentation*, 57 (2) (2001) 177-91.
25. Smith A and Thelwall M, op. cit, 363-380.
26. Jalal S K, Biswas S C and Mukhopadhyay P S, Hyperlink analysis of universities in West Bengal: a webometric study. *XXIII National seminar of IASLIC, Bose Institute, Kolkata,* 10-13 December, 2008, p.366-378
27. Mukhopadhyay P S, The calculation of Web Impact Factors for educational institutes of India: A Webometric analysis. *In. Information Management in e-Libraries,* 26-27 February 2002.
28. Li X, A review of the development and application of the Web Impact Factor, *Online Information Review,* 27(6)(2003) 407-417
29. Mukhopadhyay P S, Measuring Web Impact Factors: a webometric study based on the analysis of hyperlinks. In. *National seminar on information support for rural development, India. IASLIC*, Dec 2004.
30. Noruzi A, The Web Impact Factor: A survey of some Iranian university web sites, *Studies in Education & Psychology*, 5(2)(2005) 105-119.
31. Noruzi A, The Web Impact Factor: A Critical Review, *The Electronic Library,* 24(2006)
32. Amipour F and Payam K, Research Performance in Isfahan University of Medical Sciences: Evaluation of Scientific Productivity. *11th ISSI Conference, Madrid, Spain, 25-27, 2007*
33. Elgohary A, Arab Universities on the Web: A webometric Study, *The Electronic Library,* 26(3)(2008) 374-386
34. Jayshankar R and Ramesh Babu B, Websites in universities in Tamil Nadu: a webometric study, *Annals of library and information studies*, 56(2)(2009), 69-79
35. Ravikumar S, A webometric study of selected academic libraries in India using link analysis, *Proceedings of Fifth International Conference on Webometrics, Informetrics and Scientometrics. Dalian,* September 13-16, 2009
36. Vaughn L and Thelwall M, Scholarly use of the web: what are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology,* 54(1)(2003) 29-38
37. Mukherjee B, Measuring scholarly impact of web sites of central government universities of Uttar Pradesh and Delhi: a temporal comparison with google and MSN. *9th MANLIBNET Conference,* February 4 – 6, 2008.
38. SocSciBot 3.0. Retrieved from http://www.socscibot.wlv.ac.uk (*Accessed dated 12th Dec 2009)*
39. Pajek/Network: Programme for large network analysis. *Retrieved from http://vlado.fmf.uni-lj.si/pub/networks/pajek/ (Accessed dated 17th March 2010)*
40. Almind and Ingwersen, op. cit., 404-426
41. Ranking Web of World Universities, Methodology. *Accessed dated 19th Dec 2009 from* http://www.webometrics.info/methodology.html
42. WISER, Web Indicators for Scientific Technological and Innovation Research: A Survey of Practice, December 2005
43. Ingwersen, op. cit., 236-43
44. Smith, op. cit., 577-592
45. Jalal S K, Biswas S C and Mukhopadhyay P S, Bibliometrics to Webometrics, *Information Studies,* 15(1)(2009) 3-20
46. Thelwall, op. cit., 185-189
47. LexiURL. Retrieved from http://lexiurl.wlv.ac.uk (*Accessed dated 12th Dec 2009)*
48. Bray T, Measuring the Web. *Fifth International World Wide Web Conference*. Paris, France. May 6-10, 1996
49. Payne N, A longitudinal study of academic web links: identifying and explaining change, *University of Wolverhampton, London, UK*, 2008