

Piping “personalized Page Rank” Algorithm from the Analytic Hierarchy Process(AHP) – A Case study

K. Shanmugam¹ & Dr V Karthikeyani²

*¹Department of Computer Science, Bharathiyar University, Coimbatore, Tamilnadu, India
keshanmuga@gmail.com*

*²Department of Computer Science, Thiruvalluvar Government Arts College,
Rasipuram, Tamilnadu, India
drvkarthikeyani@gmail.com*

ABSTRACT

The core objective of any educational institution is to make a student more productive in terms of economical activity. This is simply a ripple effect towards the country’s economy. But, an extensive research work is done for ranking institutions by giving more importance for publications of research papers by the faculty members of the institutions. Everybody may not like to become a scientist. But on an average, all should survive economically. By applying one of the top 10 data mining algorithms, PageRank with the Analytic hierarchy process technique, we rank the institutions purely based on the output (alumni) instead of mixing input(students)-process(teaching-learning)-output model. Here, the candidates will arrive their own personalized ranking of institutions without depending on any national or international ranking agencies.

Keywords— Data Mining, Personalized PageRank, Institution Ranking, Analytic hierarchy process, AHP

I. INTRODUCTION

In this 21st century, not only information is overloaded but organizations, institutions, etc also. Ranking any of them is the toughest job. Hence, in this paper we aim to give a framework for ranking higher educational institutions in India. As a third nation in the world (after Nigeria and Congo as on January 2019) from suffering in extreme poverty, eradicating it is the most prioritized task. Satisfying fundamental needs are the first thing. Then comes innovation or scientific exploration of any thing. Let the innovations minimize the poverty of the people.

Still, most of the parents believe in their wards, saying “when children today in India grow up, they will be better off financially than their parents”(Pew Research Center’s Global Attitudes Survey 2017). The nation’s natural resources are becoming scarce due to the explosive growth of the population. An individual can have better wealth in participating as an entrepreneur or an employee in the world of “knowledge economy”.

An educational institution is one of the places for acquiring requisite skills to uplift a person’s economic status. But, as per 2017-18 All India Survey on Higher Education, there are around fifty

thousands of educational institutions. Every citizen should know the strengths and weaknesses of the institutions before getting an admission into them. But, human processing of the information (mostly advertisements or just for the sake of proving information) kept in their institution’s web site is tedious. Instead, they should provide web services API (application programmer interface) for programmatic processing of data for comparisons with the other institutions. Manual processing of this is also a huge task. But, all the institutions’ databases are with government. Hence, one of the nation’s duty is indexing or ranking of educational institutions after compiling their data. The authenticity of the data is also to be kept with the ranking agency but not with the institution’s. Merely, giving scoring or ranking is useless for all the stakeholders without the authenticity of the data.

As per the Survey, 2017-18, the total enrolment is 36.6 million. In that, under graduate level programmes are 79.2% (64.19 lakhs of out-turn) , and PhD is 0.5%. The total number of teachers are 12,84,755. Hence, the focus of the ranking exercise should be mainly on under graduate level in terms of employment generation. We do not have the data on “how much % of teaching faculty members are from the industries or organizations?” Whether their teaching professions are part of the work in industries or organizations? Moreover, it is not the nation’s policy.

In our model of ranking institutions, first the candidate will rank the industries by applying AHP methodology. AHP is a simple mathematical technique coming under Multi Criteria Decision Making approaches in Operations Research (OR). The methodology includes CFROI (cash flow return on investment) valuation technique as one of the criteria for assessing the performance of the industries[15].

II. A LITERATURE SURVEY

We split the survey into two categories viz., A) Statistical and Survey based techniques and B) Application of Page Rank or other Techniques

A. Statistical and Survey based Techniques

There are mainly three ranking organizations that rank the universities[10]. They are i) Academic Ranking of World Universities(ARWU) ii) Quacquarelli Symonds (QS) and iii) Times Higher Education (THE).

i) Academic Ranking of World universities or Shanghai Ranking

A world universities ranking consultancy , which entered in 2003 as the first independent organization giving ranking for global universities. This was done originally by shanghai Jiao Tong university funded by Chinese government. Their survey based findings are with some objective indicators. In one extreme (of giving weight), they see the alumni’s achievement in winning Nobel Prize or fields medals , and in another extreme, they emphasize on research output by the faculty.

ii) QS World University Rankings

These rankings are provided by British Quacquarelli Symonds annually since 2004. This ranking is also based on surveys by giving more weight to academic reputation. They use Scopus citations database for measuring research output by the faculty members. They rank global as well as regional universities.

iii) Times Higher Education(THE) World University Rankings

This rankings organization had collaborated with Thomson Reuters in 2009 but they publish rankings since 2004. Their methodologies use some performance indicators for measuring teaching, citations and research.

In the authors rank Sociology departments of the universities based on hiring patterns in the academic profession from the period of 1965 to 2007 by applying a markov model. They split the universities into two categories, “elite” group and the rest of them. They claim that there is no change in rankings between “elite” group and the rest, but change in rankings occurs within the group only.

B. Application of PageRank or other techniques

In they apply PageRank algorithm along with two other types of algorithms for producing Wikipedia Ranking of World Universities (WRWU). It ranks the universities based on the directed network of citations of the articles of 24 wikipedia language editions. In their final results, they give the geographical distribution of world-wide locations of universities along with country-wide rankings.

In the authors apply PageRank algorithm to the citation networks of research papers for giving university rankings. They look into the citation patterns among the faculty members of the universities. They adopt the approaches for five different Web of Science subjects viz., a) Dentistry, Oral Surgery and Medicine b) Business, Finance c) Information Science and Library Science d) Telecommunications and e) Veterinary Sciences.

In they attempt to give rankings for mathematics departments by taking the datasets of PhD graduates. Their work is exclusively for the positions of PhD graduates. They conclude that the PageRank algorithm is not able to give a robust ranking due to the complicated linking structures of math departments.

In also, the rankings are for PhD graduates but for Political Science Programs in the academic job market. They rank the universities by taking some of the following parameters : a) number of graduated students b) number of graduates who got the placements c) total strength of the program d) faculty size, etc. They defend that quality of the institutions means only “placement”. But the perception of the authors of this paper is that there is a chance for “recruit” the students, “hold” them for a while, “sack” them out from the back door. This probability of unethical strategy may be followed by some of the educational institutions for boosting of their score in rankings for the weight of alumni.

C. Introducing Analytical Hierarchy Process(AHP)

The AHP is a mathematical technique developed by Thomas L Saaty during 1980s as part of the decision making system[18]. “Hierarchy” means ordering of components of a system from the highest level, say root to the lowest level, say leaf. It is an inverted tree structure format. The word, “analytic” comes from the domains of mathematics as well as psychology. And finally, “process” refers to the decision making process under the hierarchical nature of the (multi criteria) decision making system. This system has a minimum of three levels as follow:

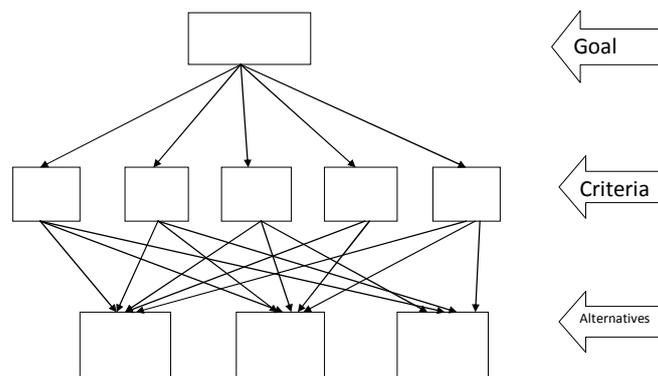


Fig 1 Adopted from

As shown above, it is decomposed into an hierarchy of criteria and alternatives for making the decision of choosing one of the alternatives. We can compare qualitative and quantitative criteria by using informed judgements for deriving priorities and weights. On the basis of relative importance of the criteria, we do pair wise comparisons. This gives us the local priorities for the alternatives with respect to each criterion. As a next step, we calculate the overall or global priority for each of the alternatives by considering all the criteria. Finally, we choose the one which has the highest priority value.

Table 1
Institutions Ranking

Sl. No.	Institution	Personalized PageRank Value	Rank
1	ins4	0.127907028	1
2	ins6	0.116284952	2
3	ins1	0.115680396	3
4	ins2	0.077305117	4
5	ins3	0.052617602	5
6	ins5	0.050748339	6

A Glance At Indian Higher Education Institutions System and Ranking Agency

A. University Grants Commission (UGC) and other autonomous educational bodies

UGC is a statutory body under Ministry of Human Resource Department, India[12]. It maintains the standards of higher education and sanctions recognition to universities, and takes care of disbursements of funds to such recognized educational institutions. Under the supervision of UGC, other autonomous bodies exist like Medical Council of India (MCI), All India Council for Technical Education(AICTE), Bar Council of India(BCI), etc.

B. Maintenance of quality in Higher Educational Institutions

National Assessment and Accreditation Council (NAAC): An independent national accreditation organization that takes care of giving accreditation for Higher Educational Institutions. It awards only Universities and Colleges but not individual programmes with an eight-grade letters : A++, A+, A, B++, B+, B, C, and D. The grades are from the highest to not accredited levels. It does the process of accreditation for the voluntary (!) educational institutions through the introspection of resources of the institutions. It is based on 34 key indicators of seven criteria.

National Board of Accreditation (NBA) : It was established by AICTE for accreditation of technical programmes (programme / branch-wise) as well as management programmes. NBA is the ranking agency for ranking academic institutions, Universities and Colleges as per the National Institutional Ranking Framework (NIRF).

In the methodology for Ranking of Academic Institutions in India , 2018, a weightage of 90% is given for the process (and the resources) and the weightage of mere 10% is given for the performance of alumni. In Software development organizations, Software Engineering Institution (SEI) [6] follows the levels for process maturity as Levels 1 through 5. Likewise, the educational institutions ranking agencies may follow the levels for the (teaching-learning) process (including resources) as the first stage, say by the NAAC / NBA . In the second stage, say, for level 5 educational institutions, the agencies may provide rankings based on the performance of output (alumni). An institution need not to wait for 20 years of existence to become a deemed-to-be-university and, recruiting faculty members producing two research papers annually on an average is a 20 days process.

Appendix Sample Runs
Fig 2 Sample Run 1

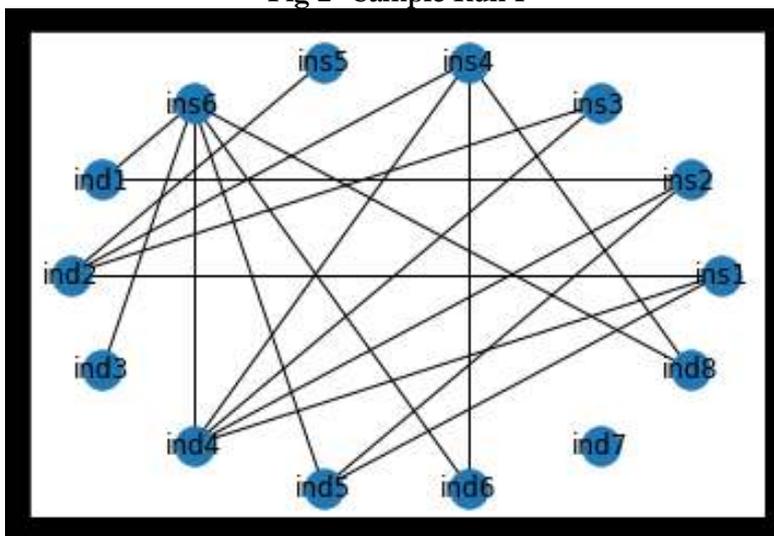


Table 2
Institutions Ranking

Sl. No.	Institution	Personalized PageRank Value	Rank
1	ins4	0.130524035	1
2	ins6	0.097754131	2
3	ins3	0.095563632	3
4	ins1	0.078026076	4
5	ins2	0.071580506	5
6	ins5	0.067095053	6

The Data Mining Methodology – Page Rank Algorithm

Page Rank algorithm is one of “Top 10 algorithms in data mining”[4][9]. It was developed by Sergey Brin and Larry Page, the google search engine founders during 1998. Basically, it is a ranking technique used for the hyper linking structure of world wide web (www). The algorithm assigns a static ranking for each web page in off-line. It works through the hyper link from a page, say p to another page, say q as a vote in favor of page q by the page p. It takes into consideration the number of votes (or in-links) a page has as well as the importance (or prestige) of the voter. Hence, while calculating the ranking for each page, the page has more prestige if it has

- i) more in-links and
- ii) how much important (or prestige) the in-linked pages have

The www forms as a graph of pages (or nodes) linked by edges (or hyper links). The page rank of the page, q is given by

$$PR(q)=(1-d)/N+d(PR(T1)/C(T1)+PR(T2)/C(T2)+...PR(Tn)/C(Tn))$$

where PR(Ti) is the PageRank of pages Ti which link to page q (in-links), C(Ti) is the number of out-links on page Ti, d is the damping factor (set between 0 and 1) , and N is the total number of all the pages on the www.

A Proposed Method

The goal here is to select the best organization by the candidate himself. The criteria are flexibility, opportunity, security, salary, retirements benefits, CFROI ranking. And the sub criteria under flexibility are location, time and work. And also another subset of sub criteria under opportunity are entrepreneurial , salary potential and top level position. There are ‘n’ number of organizations. These have to be ranked. For ranking of organizations, “ratings mode” of AHP model is applied. This forms part of

Piping “personalized Page Rank” Algorithm from the Analytic Hierarchy Process(AHP) – A Case study

the nodes in the graph and this sub graph is utilized or piped into personalized PageRank technique for ranking institutions.

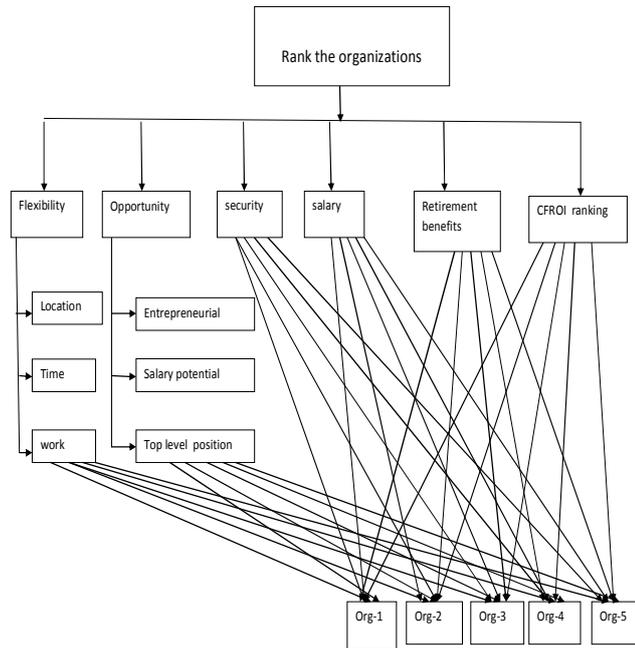


Fig 3 A modified diagram

Let V be the set of vertices (or nodes) and E be the set of edges between the vertices. Then, the network or graph, G is (V,E) . Here, the educational institutions and organizations (or industries) are treated as vertices. If an “alumnus” has got the placement in an organization, it creates a relationship (edge) between them (institution and organization). He / she may join as a faculty member in the same institution, and this forms as an edge (loop). It is not considered in PageRank algorithm as well as in our model also. But Brin and Page had removed the dangling nodes in their work but here they are treated as “unemployable” graduates. More number of such nodes lessen the prestige of the institution. If he /she may be a founder of an enterprise, he / she is treated as an employee. This “loops” (of edges) occur in the modified technique.

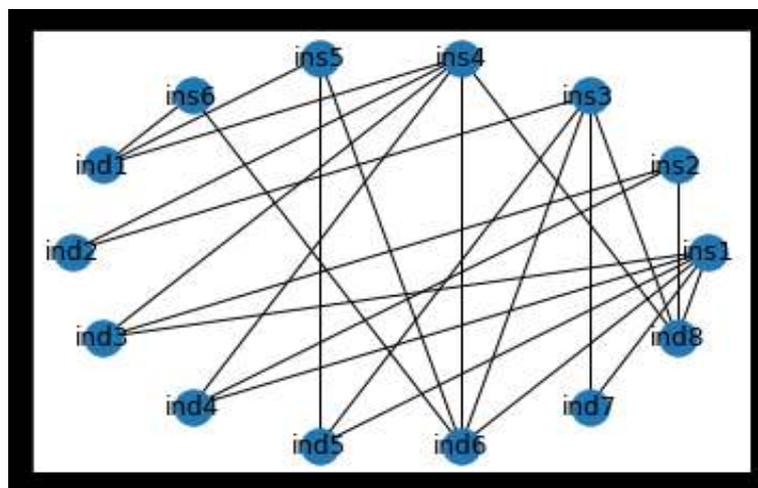


Fig 4

In all the previous works for ranking educational institutions (by PageRank algorithm), the importance or prestige due to the research work done by the faculty goes to the working institutions. Instead, it applies to the institutions, where they have studied. An institution can purchase the faculty but not the prestige or importance in terms of the algorithm.

Institutions may have its own prestige (or importance) and its maximization depends on the performance of alumni's relationships with organizations. Ordering of the organizations or industries fetches prestige or reputation to the educational institutions through the Alumni.

A Randomly Generated Dataset

As per the works of the authors in [3], the universities may follow unethical activities for boosting of their score for getting higher rankings. They split the data as i) non-corruptible parameters ii) corruptible parameters and iii) highly corruptible parameters. Even in non-corruptible parameters (of ARWU) the weight is just 30 % and it is very low. Hence, the existing dataset collected, used and ranked by the three popular world educational institutions ranking organizations viz., ARWU, QS and THE are not very impressive. For giving perfect and accurate international level of ranking of the institutions , any sampling technique may not also be followed due to trillions of graduates and billions of institutions for the hundreds of years of existence of them. Hence, we provide an analysis of a randomly generated synthetic dataset. Personalization of the algorithm is applied only to nodes having $\text{ins}(n)$, where $n=1,2,3,\dots$. The sample experiments done using Python Script are provided as Appendix – I.

III. CONCLUSION

For making policy decisions, we should have a perfect database of all the stakeholders of the educational institutions. In this paper, we strive for giving ranking to the institutions. But, finding the causes for worst performing students (or institutions) should be a national prioritized task for at least reducing unemployment problem in India. Hence, data collection, analysis, interpretation, and future courses of actions are to be initiated by the Ministry of Human Resource Department(MHRD), India for bringing our nation into the fore front in the world for reduction of Sociological Problems as well as a very good achievement in Scientific Innovations. By giving this technique as a tool for ordering industries as well as institutions, we try to give the true meaning of “autonomous” institutions. And we believe that this is the basis for “Institution of Excellence” (IOE) status.

IV. REFERENCES

1. Francesco Alessandro Massucci and Domingo Docampo ,“Measuring the academic reputation through citation networks via PageRank,” Journal of Informetrics, vol.13, pp. 185-201,2019.
2. José Lages, Antoine PattDima and L. Shepelyansky, “Wikipedia Ranking of World Universities”, The European Physical Journal B, 89(3):69, 2016
3. Maxime Mussard and Alex Pappachen James, “Engineering the Global University Rankings: Gold Standards, Limitations and Implications”, IEEE Access, vol. 6, pp. 6765 – 6776, 2018
4. Wu, X., Kumar, V., Ross Quinlan, J. et al., “Top 10 algorithms in data mining”, Knowledge and Information Systems, vol. 14, pp. 1-37, 2008
5. Jacob Mincer , “Investment in Human Capital and Personal Income Distribution”, Journal of Political Economy, vol. 66, pp. 281-302, 1958
6. W.S. Humphrey, “Characterizing the software process: a maturity framework”, IEEE Software, vol. 5, pp. 73-79, 1988
7. Teresa Ratashak , “Ranking Math Departments Using the Google Page Rank Algorithm”, , University of Oklahoma, Honors Math Thesis, 2014
8. Benjamin M. Schmidt and Matthew M. Chingos , “Ranking Doctoral Programs by Placement: A New Method”, Political Science & Politics, pp. 523-529, 2007

Piping “personalized Page Rank” Algorithm from the Analytic Hierarchy Process(AHP) – A Case study

9. Sergey Brin and Lawrence Page , “The anatomy of a large-scale hyper textual Web search engine”, Computer Networks and ISDN Systems, pp. 107-117, 1998
10. AP Pavel , “Global University Rankings - A Comparative Analysis”, Procedia Economics and Finance, vol. 26, pp. 54-63, 2015
11. Weakliem, D.L., Gauchat, G. & Wright, B.R.E. , “Sociological Stratification: Change and Continuity in the Distribution of Departmental Prestige, 1965–2007”, vol. 43, pp. 310-327, 2012
12. ugc.ac.in
13. nirfindia.org
14. naac.gov.in
15. Madden, B.J., CFROI Valuation: A Total System Approach to Valuing the Firm, Butterworth-Heinemann, Oxford, 1999
16. T.L. Saaty, Decision making with the analytic hierarchy process Int. J. Serv. Sci., 1 (1) , pp. 83-98, 2008
17. T.L. Saaty, L.G. Vargas, Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, vol. 175, Springer, 2012
18. T. Saaty, Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in a Complex World, Belmont, CA:Wadsworth, 1982