Available online at www.sciencedirect.com
SciVerse ScienceDirect
ELSEVIER
Procedia - Social and Behavioral Sciences 73 (2013) 518 – 525
Procedia
Social and Behavioral Sciences

The 2nd International Conference on Integrated Information

Institutional Research Management using an Integrated Information System

Tsolakidis Anastasios[a,b], Sgouropoulou Cleo[a], Papageorgiou Effie[a], Terraz Olivier[b], Miaoulis George[a,b]

[a]*Technological Education Institute of Athens, Ag.Spyridonos St., 122 10 Egaleo, Greece Tel: (+30) 210 53 85 826*
Email: {atsolakid, csgouro, efipapag, gmiaoul}@teiath.gr
[b]*Université de Limoges, Laboratoire XLIM, CNRS, UMR 6172 83, rue d'Isle, 87000 Limoges France*
Email: {olivier.terrazr, georges.miaoulis}@xlim.fr

Abstract

Research constitutes a fundamental activity within Higher Education and, for many institutions, comprises a major revenue income stream. Moreover, research increasingly becomes dependent on information systems to manage the complex information flows that it creates. This study introduces a web based software system built on Java technologies that supports institutional research management. The system provides decision support mechanisms using graph metrics in combination with data envelopment analysis as a method for efficiency measurement. Comparing the R&D outcomes of academic units with the dynamics of the collaboration patterns extracted from graphs, the developed system enables research managers to evaluate specific criteria and correlate strategic goals with research performance.

Selection and peer-review under responsibility of The 2nd International Conference on Integrated Information

Keywords: research management; graph metrics; DEA;information system; visual analytics;social network analysis; info graphics

1. Introduction

Research and Development (R&D) systems in Higher Education Institutes (HEI) play an important role in generating economic and social growth. Due to the importance of research administration within the academic context, in this paper introduces an integrated Information System which has been developed in order to support the management of institutional research resources. The main objective of the system is to characterize and assess the collaboration patterns among the faculty members of an academic unit in the perspective of the research policy of the Institute. The assessment focuses on R&D efficiency and effectiveness ,based upon the quality criteria the user of the system sets, providing as well the correlation among research performance and research collaboration. The case study presented in the paper builds on the research activities of the faculty members of a Greek HEI within the period 2009-2010. The proposed Information System implements graph analysis methods

* Corresponding author: Tsolakidis Anastasios Tel.: +30 210 5385826.
E-mail address: atsolakid@teiath.gr

Selection and peer-review under responsibility of The 2nd International Conference on Integrated Information
doi:10.1016/j.sbspro.2013.02.085

based on the collaboration network created from the co-authoring of research publications. In order to build such a network we use data harvested from Scopus. The main idea is to provide an approach towards the measurement of the cooperation among the faculty members of the academic unit using the collaboration graph and evaluate their research performance through the Data Envelopment Analysis (DEA) methodology [0] . By means of it, we aim to answer questions such as: “How active is the given author?”, “Which author could be considered as a research hub?” and, “How is collaboration correlated with performance?”. Our approach, compared to other methods presented in [0,0], combines info-graphics with data mining. The developed system provides decision support to the user leveraging methods which analyze the data and discover further hidden patterns of collaboration among the HEI faculty members and integrating visual representation techniques. More specifically, the system implements analysis of research collaboration networks, Cluster Analysis based on graph metrics and research performance measures using DEA.

2. Related Work

The decision-making process in HEIs should be planned and resolved in a comprehensive, reliable, and transparent manner [0] in order to improve the quality of education and research. Decision Support Systems (DSS) are designed to help decision makers compile useful information from a combination of raw data to identify and solve problems and make decisions. Various academic DSSs have been proposed for handling issues, such as resource allocation [0,0,0] workload management [0], course scheduling, admission policy [0], advising [0] , and strategic planning[0]. The proposed Information System has been developed in order to manage the research activities of a Higher Education Institute from the perspective of social network analysis. The main objective is to develop a system which represents the research collaboration networks and the research performance of its faculty members. In addition, the system provides the correlation between R&D activities and performance. Relevant studies have emphasized the importance of collaboration, in the context of national R&D policy [0,0] based on the number of publications and the impact factor of papers as output measures. Furthermore, the R&D productivity is measured through DEA, being one of the most common methodologies [0,0] which uses linear programming for measuring and benchmarking the relative efficiency of different Decision-Making Units (DMUs) with multiple input and output.

3. System Analysis

In the following paragraphs we present a decision support system for the assessment of research activities of the faculty members within HEIs, using visual analytics. The system analysis and respective implementation is to be considered on the basis of four hierarchical and complementary levels, namely *Data Collection*, *Data Processing*, *Data Mining* and *Interactive Decision Making*, as those described in Figure 1. The main aim is the integration of data mining methods as graph analytics, clustering and classification, encompassed by the Data Envelopment Analysis method, using visual representations.

3.1. Data Collection

Concerning data collection we have decided to integrate with Elsevier’s large database (Scopus) [0] as our primary source. By using Scopus, we were able to construct a database that contains all faculty publications, which were co-authored by faculty members. The co-authoring network created is the "network" that we use for our Analysis. Furthermore, we combined publication data with R&D annual reports in order to analyse the structure of the faculty co-authorship network and calculate centrality measures and research efficiency.

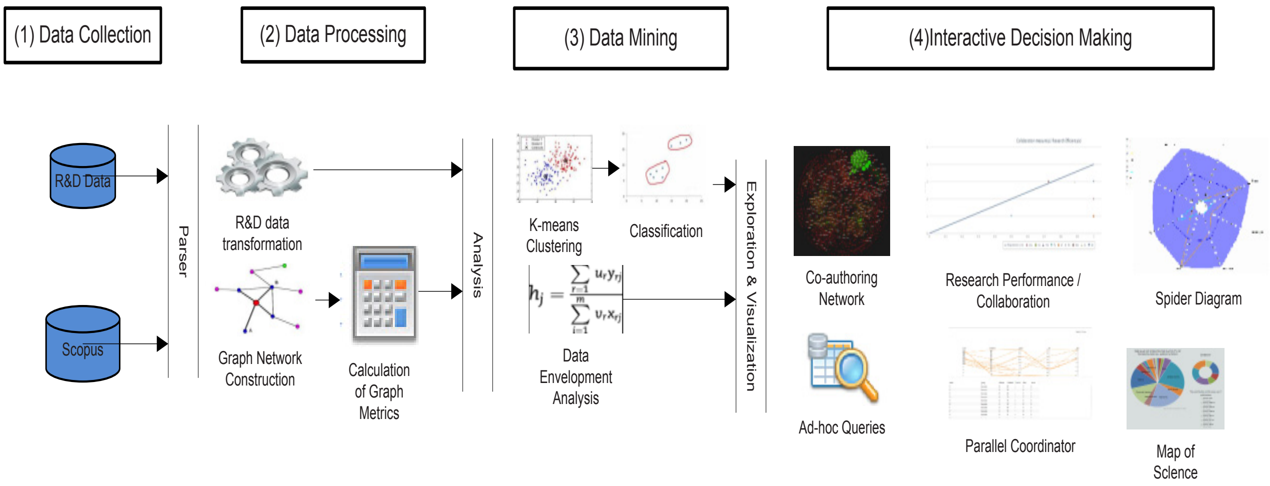


Figure 1. Information System for Institutional Research Management

3.2. Data Processing and Graph Analysis

The system implementation builds upon graph networks and the relationships that may exist among the academic researchers based on the co-authoring of research papers. In order to create the network we set the authors as nodes. Their collaboration in a publication is denoted by the network edges. Hence, we construct a binary $N*N$ matrix where N is the number of the authors (nodes).Our system uses the JUNG (Java Universal Network / Graph) Framework [0], a software library that provides a common and extendible language analysis of data that can be represented as a graph or network. On this basis, our application implements a number of important measures as centralities, PageRank and HITS (hubs and authorities).

3.3. Data Mining

3.3.1. Collaboration Measure

In subsequence to the calculation of the graph metrics for the whole topology we provide the user with the ability to classify the authors into groups based on clusters created using the K-means algorithm. After the execution of the algorithm the user may view the clusters and classify those with the highest degree as the best and with the lowest degree as the worst group. The range of the values depends on the number of the clusters. For example, for four clusters we have four values. The decision for using the K-means cluster analysis method [0] is because it aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The attributes of the instances for clustering are the graph metrics which show the degree of collaboration.

3.3.2. Efficiency Measure

Data Envelopment Analysis (DEA) is a deterministic frontier technique, whose aim is to model the “production process” of homogenous organizations, to derive an “efficiency frontier”. It is based on a linear programming technique, introduced by Charnes et al. [0]. In a DEA model, the notion of efficiency is that of “technical efficiency”, defined as the relative ability of each Decision Making Unit (DMU) in producing outputs

[0]. The most important notion is the word "relative" which means that we compare a unit with any other homogeneous unit. DEA could be represented by a linear programming technique where each DMU tries to maximize the efficiency ratio (output over inputs) choosing the best set of weights. By looking at the associated weights as those derived by the DEA, it is possible to understand how the inputs and outputs contribute to the overall performance. In order to implement the DEA technique the user uploads a .xls file, each line of which contains the units and each column the output or input values.

3.4. Data Visualization

3.4.1. Co-authoring Graph

A co-authoring (or co-authorship) network is the network created on the basis of the collaboration among the faculty members for the publication of a research paper [0]. Figure 2 illustrates such a co-authoring network, where nodes represent the authors and edges the collaboration (co-authoring activity) among them. Each of the edges has been assigned with a value denoting the number of citations for the paper. The diameter of the nodes and their color depends on the number of publications the authors have submitted. In our implementation the user may select a node and view only the authors that have links to the selected author. The appearance of the graph is based on the criteria (centralities, PageRank and HITS) that the user has selected. Different criteria provide different graph networks, thus resulting in different nodes' colour and size.

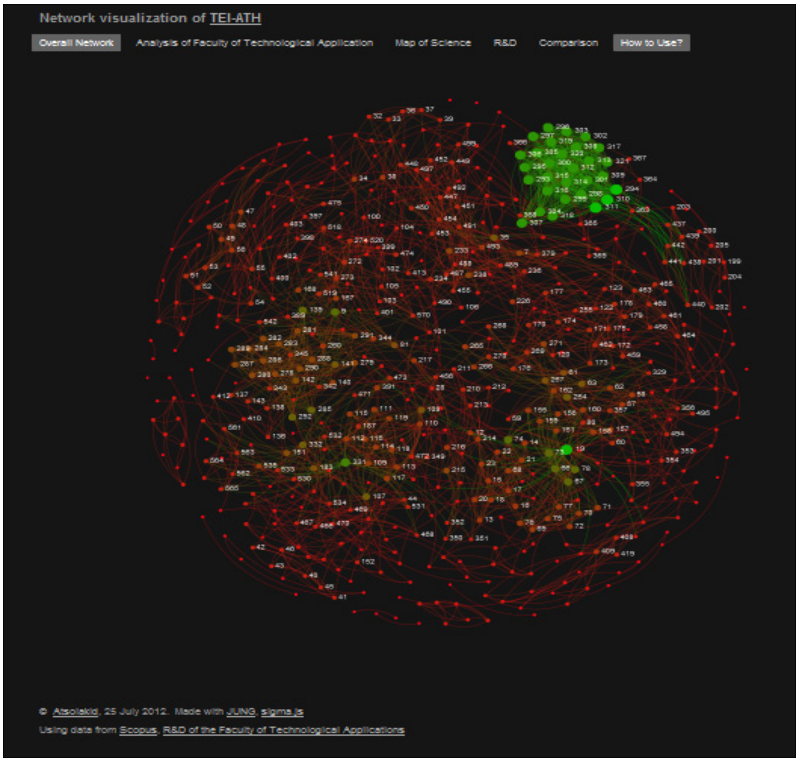


Figure 2. Co-authoring Network

3.4.2. Collaboration vs Efficiency

The main objective of our system is to visualize different data mining approaches in order to help decision makers to get an accurate view of the examined sectors. In this direction we develop visualizations that provide the coloration among different metrics. For example, using the application the user may observe the correlation among collaboration metrics and research efficiency of the examined faculty members.

In order to use this functionality the user has to follow three steps:

1. Initially the user executes a K-means cluster analysis based on the graph metrics values that have been calculated at the time of the graph construction. These values have been selected because all of them represent the collaboration efficiency of a node.
2. Then, the user classifies the clusters using as High Performance (4), Normal Performance (3), Low Performance (2), Very Low Performance (1).
3. Finally, the user executes the research performance measure based on the R&D data that he/she has uploaded. So the user launches the application (java-based) and gets the scores.

After these steps, using the graph representation functionality, the user observes the coloration among collaboration and research performance. In Figure 3 we may see 9 instances represented with different colors and different shapes. For example, we observe that there is an instance with the highest degree of collaborations as well as the highest value of research performance.

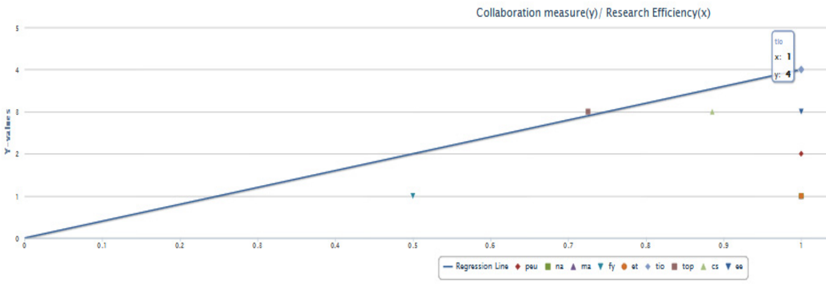


Figure 3. Collaboration / Research Efficiency

3.4.3. Parallel Coordinators

One of the most common research evaluation criteria is to provide the most qualified researcher among the faculty members. Using the parallel coordinator visualization method [0] the user is able to apply a set of criteria depending on his objectives. Figure 4 illustrates the parallel coordinator visualization which uses, as case study, the faculty members of a Department of Informatics of a Greek HEI. The parallel lines represent the criteria that the user has set based on the .xls file he has uploaded. The criteria of our example are the h-index, the post(Professor, Assistant Professor,..), the projects, the number of publications and the research field of the faculty members. Using this representation the user has the ability to set filters manually and select only those data with values within specific, desired ranges. Hence, in Figure 4 the user may view all the 25 authors of the department. Figure 5 shows the application results on the basis of the h- index criterion and, specifically, the authors with h-index among 4 and 9.

Using parallel coordinators, the user could set priorities by selecting the range of values for the examined fields. So the user defines the priorities by observing the overall quality of all the faculty members. We thought that it is of high importance for a decision maker to take decisions in that way, because at any time he could change the ranges of the values or the order of the filters so as to get different results and different decisions.

The types of queries that our system supports are:

Linear

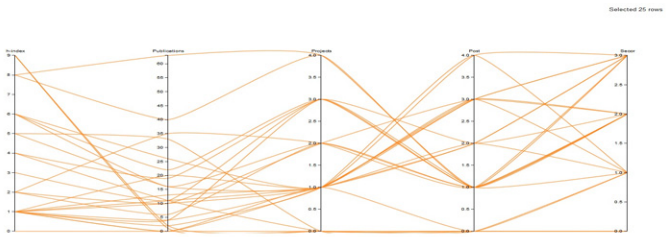
e.g. "Find all the authors with their publications"

Filtering

e.g. "Find all the authors who have more than 10 publications"

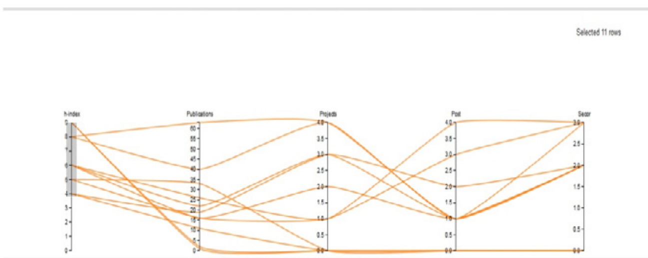
Structured

e.g. "Find all the authors and the corresponding co-authors".



name	group	h-index	Publicati...	Projects	Post	Secor
12	Informatics	8	63	4	1	2
17	Informatics	2	35	2	1	3
23	Informatics	6	26	1	4	3
18	Informatics	8	40	4	1	3
14	Informatics	4	19	3	2	2
10	Informatics	6	16	2	1	2
58	Informatics	5	33	0	0	0
5	Informatics	1	20	1	3	1
3	Informatics	1	11	1	2	1
11	Informatics	6	22	3	1	2
15	Informatics	2	10	2	3	2
21	Informatics	5	16	1	3	3
16	Informatics	1	10	1	3	2
20	Informatics	1	13	1	2	3
13	Informatics	3	11	3	1	2
33	Informatics	4	11	0	0	0
9	Informatics	2	6	2	1	2
19	Informatics	1	4	1	2	3
8	Informatics	1	2	1	4	1
26	Informatics	9	1	0	0	0
27	Informatics	9	2	0	0	0
4	Informatics	1	0	1	0	1
22	Informatics	1	0	1	3	3
2	Informatics	0	0	0	0	1
1	Informatics	1	4	3	1	1

Figure 4. Parallel coordinator for all the faculty members without any criteria



name	group	h-index	Publicati...	Projects	Post	Secor
12	Informatics	8	63	4	1	2
23	Informatics	6	26	1	4	3
18	Informatics	8	40	4	1	3
14	Informatics	4	19	3	2	2
10	Informatics	6	16	2	1	2
58	Informatics	5	33	0	0	0
11	Informatics	6	22	3	1	2
21	Informatics	5	16	1	3	3
33	Informatics	4	11	0	0	0

Figure 5 Show all the faculty members with h-index value besides 4-9

3.4.4. Spider Diagram

The Spider diagram has been selected as a visualization tool to be used because it is the most suitable for the representation of evaluated instances because it provides a sense of the overall performance in comparison with the selected instances. This graphic attempts to visualize the R&D performance. The outer radius of the ring shows the maximum of the indicators and the inner radius the minimum. The thick blue line shows the median. In Figure 6 eight indicators have been selected to show the department's R&D performance. In this example we compare 2 departments, selected through the checkboxes on the left and observe the corresponding (color) line in the graph.

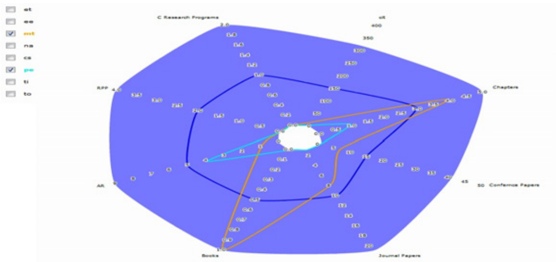


Figure 6. Spider Diagram

4. Conclusion

The proposed Information System is illustrated by evaluating research activities through a biblio metric analysis. The results are provided through “visual analysis” which enable the user to characterise and assess the collaboration patterns among the faculty members of an academic unit in the perspective of the research policy of the Institute. Future work falls under the prospective to add more data mining algorithms because the model of visualization tool that we have initially developed has capability to be extensible in order to support for adding algorithms and various datasets.

Acknowledgements

This work has been funded by the National Project “MODIP of TEI of Athens” – NRSF 2007-2013.

References

[1] Charnes A, Cooper WW, Rhodes E . ‘Measuring the Efficiency of Decision Making Units’. European Journal of Operational Research 1978;2:429-444.

[2] Suraj, Z., Grochowalski, P.: Patterns of Collaborations in Rough Set Research. Studies in Fuzziness and Soft Computing, vol. 224, pp. 79–92. Springer, Heidelberg (2008)

[3] Suraj, Z., Grochowalski, P., Lew, _L.: Pawlak Collaboration Graph and Its Properties. In: Kuznetsov, S.O., Slezak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS, vol. 6743, pp. 365–368. Springer, Heidelberg (2011)

[4] Shimizu T., de Carvalho M.M., Laurindo F.J.B., Strategic Alignment Process and Decision Support Systems: Theory and Case Studies, Idea Group Inc., 2006

[5] Burke E.K., Kendall G., Search Methodologies. Introductory Tutorials in Optimization and Decision Support Techniques, Springer Science and Business Media, 2005
[6] Chen Z., Computational Intelligence for Decision Support, CRC Press LLC, 2000
[7] Doumpos M., Zopounidis C., Multicriteria Decision Aid Classification Methods, Kluwer Academic Publishers, 2004
[8] Forgionne G.A., An Architecture for the Integration of Decision Making Support Functionalities, Decision Making Support Systems: Achievements, Trends and Challenges for the New Decade, Idea Group Inc., 2003 Gago P., Santos M.F., Closed Loop Knowledge Discovery for Decision Support in Intensive Care Medicine, Recent Advances in Computers, 13th WSEAS International Conference on Computers, WSEAS Press, p. 447-452
[9] Grünig R., Kühn R., Successful Decision-making. A Systematic Approach to Complex Problems, Springer-Verlag Berlin Heidelberg, 2005
[10] Gupta J., Forgionne G.A., Mora M.T., Intelligent Decision-making Support Systems. Foundations, Applications and Challenges, Springer-Verlag; 2006
[11] Hashim F., Alam G.M., Siraj S., Ensuring participatory based decision-making practice in Higher Education through E-management: A faculty initiative, Recent Advances in E-Activities, Information, Security and Privacy, WSEAS Press
[12] Defazio, D., Lockett, A., & Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. Research Policy, 38(2), 293–305. doi: 10.1016/j.respol.2008.11.008.
[13] Mote, J. E., Jordan, G., Hage, J., & Whitestone, Y. (2007). New directions in the use of network analysis in research and product development evaluation. Research Evaluation, 16(3), 191–203.
[14] Kao, C. & Hung, H. (2008). Efficiency Analysis of University Departments: An Empirical Study. Omega, 36(4), 653–664. doi:10.1016/j.omega.2006.02.003
[15] Tommaso Agasisti, Giuseppe Catalano, Paolo Landoni and Roberto Verganti . Evaluating the performance of academic departments: an analysis of research-related output efficiency. Research Evaluation (2012) 21(1): 2-14.doi: 10.1093/reseval/rvr001
[16] Scopus
[17] Universal Network/Graph Framework
[18] A hybridized k-means clustering approach for high dimensional dataset- By Rajashree Dash, Debahuti Mishra, Amiya Kumar Rath, Milu Acharya, Orissa, International Journal of Engineering, Science and technology, Vol 2, No. 2, 2010, pp. 59-66
[19] Tsolakidis A., Sgouropoulou C., Xydas I., Terraz O., Miaoulis G., Academic research policy-making and evaluation using graph visualisation. PCI 2011 , art. no. 6065059 , pp. 28-32
[20] Inselberg, A. (1998) Visual Data Mining with Parallel Coordinates. Computational Statistics, 13, 47–63.