Academic evaluation and research policy decision making using graph visualisation

Anastasios Tsolakidis[1,2], Cleo Sgouropoulou[2], Ioannis Xydas[2], Olivier Terraz[1],.Georgios Miaoulis[1,2]

XLIM – UMR 6172 – CNRS,
Department of Mathematics and Computer Sciences
123, avenue Albert Thomas - 87060 LIMOGES CEDEX, FRANCE,
Email: olivier.terraz@xlim.fr, georges.miaoulis@xlim.fr,

Technological Education Institute of Athens
Department of Informatics
Ag.Spyridonos St., 122 10 Egaleo, GREECE
Email: atsolakid@teiath.gr, csgouro@teiath.gr, yxydas@teiath.gr, gmiaoul@teiath.gr,

Abstract

The educational institutions need to have a clear mission, comprehensive vision and attainable objectives that can be interpreted and analyzed by using terms of strategic planning. Our work is part of prototype system that supports academic evaluation and decision-making strategies concerning the research policy of an institute, using visual analytics. It focuses on the analysis and visualization of the relationships that exist between authors, scientific publications and research fields, within an academic unit. Using, initially, the research field as a filter we create a co – authoring network of authors connected to each other through weighted associations. Weight values are produced by applying more than one algorithm to key attributes of the publications (citation, publisher, etc.) and characterize the connections (edges) of the network. The graph nodes represent authors (researchers) with their attributes. Our data is enriched by analysis and metrics of the graph. The online use of successive vectors of thresholds in the characteristics of the graph enables users to get a “visual multi-criteria analysis” for the validation of the main hypotheses of institutional research policies and the support of further decision making.

Keywords

Visualization, Co-authoring, Knowledge discovery, Decision support systems, Research policy

Introduction

Educational institutions increasingly need to assess and enhance their activities, in order to provide a balance of tangible and intangible assets, and to measure future capability as well as past performance. Research effectiveness is, among others, an important measure of the quality assurance process in Higher Education.

A co-authoring network within an academic unit is a set of authors / researchers, each of which has connections to some or all of the others. Through our analysis we can visualize this network using graph representation, where authors are represented by nodes and their connections by edges. A connection between two scientists is established through their co-authorship in one or more scientific publications. Both the nodes and edges can be defined in different ways depending on the criteria of interest. A well-structured network could provide the most significant features in a directly perceptible way. For example, such a representation could provide answers to specific research collaboration related questions, i.e. “are the relationships (edges) among the authors enough?”; “are there any co-authoring communities?”; “how many edges depart from / end to a specific author?”, etc. Our research demonstrates that there is not a unique, optimal way for representing the data supporting responses to the above questions. However, we can transform our graph according to the information and relationships we wish to represent. On this basis, we depict the relations that exist between research authors in a specific domain, in a way that users may observe existing communities of authors with stronger links than others and the thematic areas of their publications. In order to provide a weight to these relations we count the number of publications of each of the authors in collaboration to some, or all of the others. Subsequently, we use community detection algorithms in order to generate the representation of existing research communities. We examine the thematic areas of their publications and compare them with the keywords of the papers and the authors that belong to those areas. Finally, we set a research policy that is described by specific performance criteria and create the appropriate representation graph.

In this article we start our discussion with the basics of the academic evaluation process and the importance of research activity for the enhancement of institutional knowledge and innovation. Within this context, our work leverages information visualization techniques as the basis for providing intuitive and interactive decision support to institutional research policy making. We focus on and describe co-authorship networks as research information-rich sources through which we may observe emerging clusters and communities. Moreover, we propose an ontology based software system architecture for supporting research policy decision making though graph visualization. Finally, we outline the main conclusions of our work and discuss future directions.

The research indicators in academic evaluation

From an institutional perspective, academic evaluation is the systematic and on-going process of collecting, analysing and interpreting data related to specific indicators and objectives, developed to support the mission and purpose of the institution.

Educational institutions need to have a clear mission, comprehensive vision and attainable objectives [1]. The academic evaluation process is built around the articulation and development of measurable criteria and performance indicators, providing both internal and external stakeholders with the opportunity to understand which are the objectives of the institution, how does it pursue their achievement, how does it know that it has succeeded in achieving these objectives, what adjustments would ensure success and how well the achievement of the objectives is linked to specific strategies.

In order to implement evaluation within a Greek academic unit, the first important step has been to design a prototype process that integrates and harmonises institutional strategies and state defined requirements based on the Hellenic Quality Assurance Agency for Higher Education (HQAA) framework and the related areas of quality criteria and standards, namely quality of teaching, quality of research, quality of curricula and quality of services and infrastructure [2].

The overall evaluation exercise has been useful, in the sense that it helped realizing strengths and weaknesses (and primarily the reasons for them) through a systematic analysis and that it served as a training process in quality issues. In addition, it has provided an important basis for the forthcoming institution-wide implementation.

Undoubtedly, the quality research function of an institution remains a prime source of knowledge and innovation. The formulation and implementation of institutional research policies constitutes a strategic priority for academic units, having as principal targets the enhancement of research-based teaching and learning and the forging of links with intra and inter-institutional research communities.

In support of this process, our work aspires to provide a solid instrument for institutional research policy makers, enabling them to set strategic objectives and, through visual observation and analysis of the research population's activities, to make decisions for future directions and priorities.

Related work

The mapping of the scientific research activities has become a well-established visualization routine in the field of information science. Small and Griffith [3] established the principles of co-citation analysis, introducing a methodology for observing the specialists' activity in a particular domain. In general, each publication is associated with the author(s) who wrote it, the references it cites, the conference or journal in which it was published, and the keywords / terms that appear in it. The Paper Box model [4] suggests a model using information about: co-authorship, bibliographic coupling, paper co-occurrence and term co-occurrence, as well as rules to be applied (such as Bradford's Law) [4].

Various visualization techniques are used in order to represent the co-authoring networks based on the authors or the documents that they publish. The most commonly used method is the node link network visualization, developed in tools such as CiteSpace [5] and Pajek [6]. The visual analysis of these data is the most critical aspect and a lot of techniques have been developed in order to deal with this challenge. Some of the field researchers are based on network degrees and community detection algorithms, while others, on the layout and the kind of the representation [7] used.

Community detection algorithms such as leading eigenvector, walktrap, edge betweenness and spinglass [8] are used in order to assign structural communities to authors. Moreover, researchers use different kinds of detection algorithms, enabling user to extract the information perceived as more important.

Regarding the field of research policies, authors in [9] suggest a methodology in order to achieve high level of integration in research (as a part of the research policy) and also the way that we could evaluate the success or failure of a research policy.

Interactive decision policy making

The related literature study clearly indicates co-authorship networks as research information-rich sources. Co-authorship networks are used in order to determine the scientific collaborations, the strength of researchers and the thematic areas of their interest. For the purposes of this article we created a network of scientists, where a connection between two authors is established when they have a common publication.

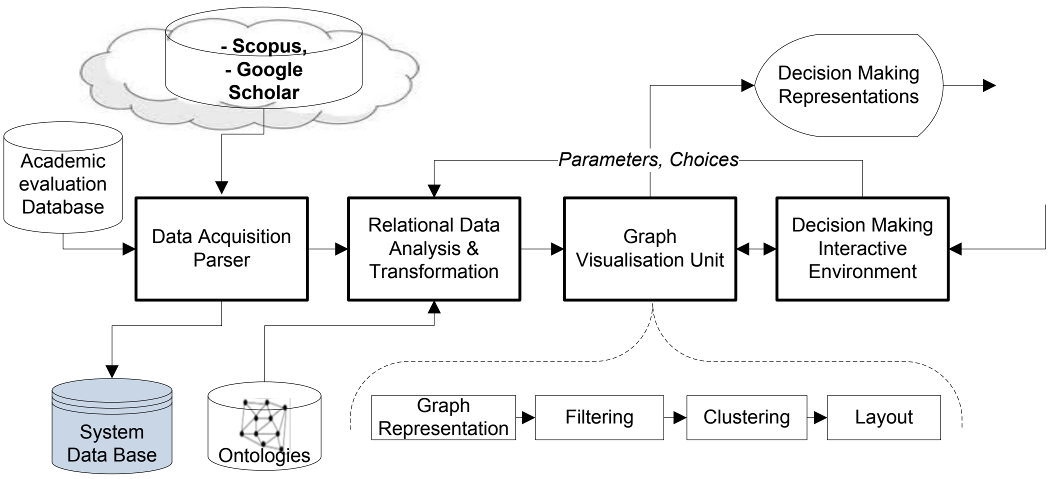

Figure 1 System Architecture

The data that we used represent the scientific publication activity of the 88 members of the academic staff of the Department of Informatics, at the Technological Educational Institute of Athens. The academic staff includes professors and fellows. Based on these data, we have created a graph representation, where authors constitute the nodes of the graph and the edges represent the collaborations between them. The authorship data have been retrieved by the department's academic evaluation database, whilst for the keywords and thematic areas we used a web service for harvesting the Scopus and Google Scholar library.

In order to visualize the institution's research activities based on co-citation relationships, we have used the existing clustering coefficient and modularity algorithms [10,11] , due to our attempt to observe the clusters and the communities that may exist. In addition, we have enriched the framework for graph visualization used in the PGA prototype [12]. Our data are represented by means of a Java-based interactive prototype using the Gephi visualization tool [13]. Gephi was selected because of its ability to represent high quality graphs in networks up to 50K nodes and 500K edges.

The overall architecture of our system is depicted in Figure 1. The Graph Visualization Unit integrates the aforementioned algorithms, techniques and visualization tools. Compared to other research activity visualization tools [5, 6,12], our system supports the incorporation of ontologies, modelling (existing or desired) institutional research aspects (i.e. areas, rules, objectives). This feature enables policy makers to set target policies and monitor current and future performance towards the specified directions. Moreover, the user could be provided with additional kinds of representations in order to be further supported towards more accurate decisions.

In our attempt to represent the most publication-active authors in the network, we decided to apply a weight to the link that connects the scientists through their common publication. Initially, the value of an edge is set to one. Each time a pair of authors establishes a new collaboration, this value is increased.

On this basis, the final network comprises all authors (nodes) that belong to the department and the connections (edges) that represent the collaboration among them. Each of the edges may have a specific value, equal to the weight of the link between the authors.

Clustering Coefficient

Clustering coefficient [9] is a measure of degree to which nodes in a graph tend to cluster together. It shows how well connected the neighbourhood of the node is. If the clustering coefficient is 1 then it means that the neighbourhood is fully connected, and if it is 0 then it means that there are no connections in the neighbourhood.

From this representation we can observe which authors tend to cluster together. The size and the colour of the nodes depend on their clustering values.

Community Detection -Louvain method

In order to detect the communities that may exist in our network we use the Louvain method [10]. This method consists of two phases. Initially, it looks for "small" communities by optimizing modularity in a local way. Secondly, it aggregates nodes of the same community and builds a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained.

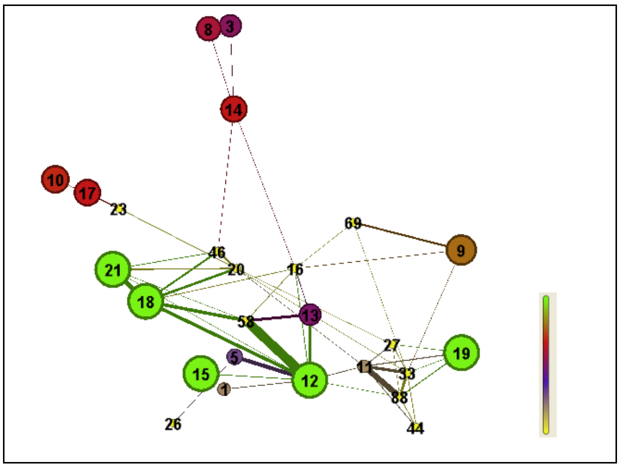

Figure 2 Clustering Coefficient

In this graph (figure 3) we observe that there are 5 different colour groups. In each of these groups the member authors are represented. The graph may illustrate big communities with more authors than others, communities whose authors have a large number of publications and communities that seem of low interest because their members do not have enough publications.

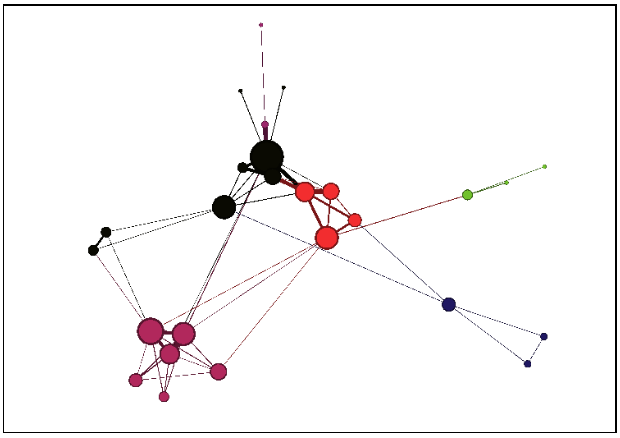

Figure 3 Louvain Method

Interactive Visual Support for Academic Research Policy Making

The setting of a research policy in an academic unit is driven by a number of significant parameters. Our system equips the user with all the necessary representation tools in order to specify and validate a research policy. Supposing that an academic unit decides to change its research policy and set new priorities for activity on the promotion of innovation in an interdisciplinary research area (e.g. learning design, intelligent systems and information visualization).

Our system, through the alternative visualization representations of researchers' activity in specific thematic areas, allows the user to observe which area is heavily represented, if the newly targeted areas are individually active, which are the most contributing researchers in these areas, if there are any existing links among authors that could generate the necessary dynamics for the evolution of the targeted interdisciplinary area, etc.

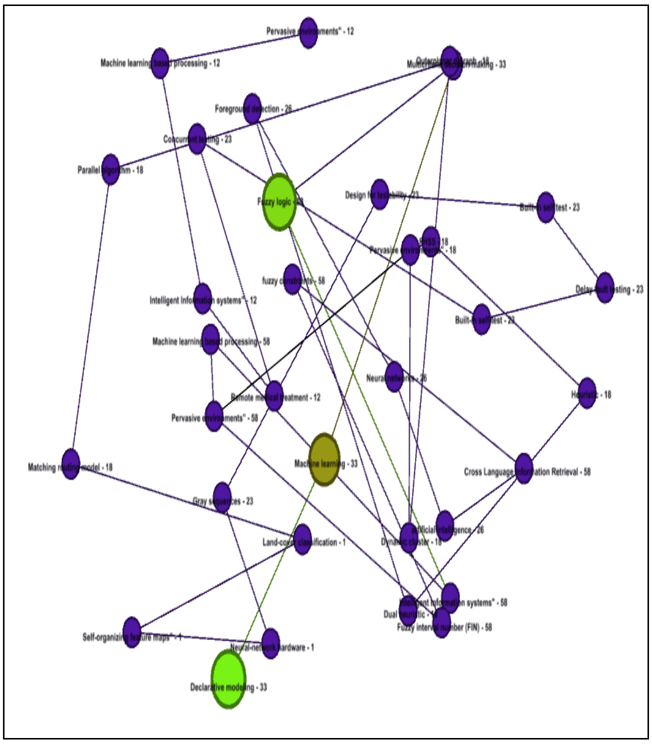

Figure 4 Keywords for a thematic area

This functionality provides policy-makers with the opportunity to efficiently validate their hypotheses, specify concrete actions towards their implementation, or even decide to discard them in case they lack necessary foundation. Another type of functionality our system provides is for the user to select a field of interest (thematic area) and subsequently, the keywords that he/she wishes to investigate on.

A graph is produced based on the specified criteria. From the topology of the graph the user accesses all the information related to the most active scientists. Let us, for example, suppose that the department wants to set as research policy for the next period the area of artificial intelligence and wishes to find which scientists have research activities in this area. The user sets as a filter the thematic area of Artificial Intelligent (figure 4) and gets the keywords that the scientists have set in their publications pertaining to this area. At the graph of figure 4 the nodes represent the keywords and each of the nodes corresponds to a specific author (the number next to the keyword). The size of the nodes depends on the number that each one of the keywords was included in the papers.

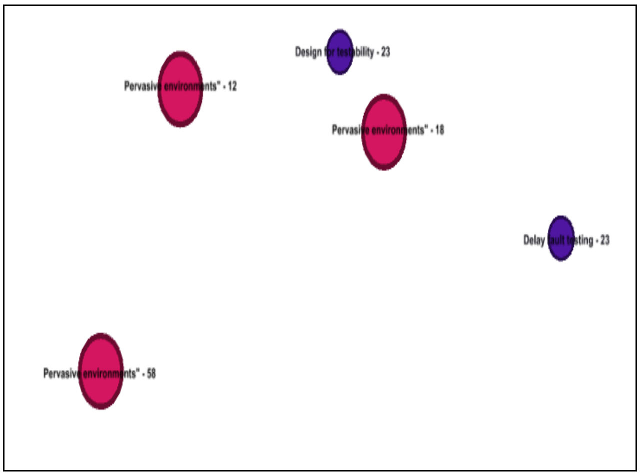

Figure 5 Specific Keywords for a thematic area

A next step (figure 5) would be to view the keywords of the most interesting (high weight value) or to set the keywords of interest to recognize the authors with publications relevant to these areas. In both representations we observe the keywords and the number of the authors. So if the department wants to set as policy for the next period the area of artificial intelligence they view all scientists that have publications in this area and also they can observe the most frequently used keywords.

Conclusions – Further Work

The application field of the described methodology has been a Greek Higher Education academic unit of the technological sector, which has already applied conventional academic evaluation processes. Three independent stakeholders used this prototype of interactive tools in order to assess the proposed methodology and tools. The stakeholders used the database of the given system with the files of the department's research and teaching staff, the scientific publications and also a research fields ontology, associated with their attributes. Our system provides the structure of the potential research groups of this department in combination with the fields of research under development.

The two most significant functions of the system are (a) on one hand, the possibility of hypothesis-based investigation for composite research fields, meaning research areas with a combination of more fields or "keywords" and the ability to represent visually potential structures of research groups for this department; (b) on the other hand, the possibility of direct and interactive visualization of sensitivity analysis results for the characteristics of the systems entities such as research staff and publications, projected according to «what if» development scenarios of the given research policy choices.

Throughout the experiments we used different types of structure's graphs visualization and their values, by altering the various colours and sizes on the level of the nodes and the arcs. The emphasis was given to the quality of visual perception and the observability of represented phenomena.

The first results of this prototype experimentation, using the aforementioned interactive visual utilities by those three independent stakeholders, proved the feasibility and the utility of such a system.

Future work falls under the prospective to improve the main functionalities of the system. In addition, we will have the opportunity to enhance the system evaluation in a more systematic way, by broader sets of data (included several departments) and by engaging a number of decision makers in the field.

References

1. Kerstin V. Siakas, The Design of a Quality Assurance System in Higher Education – Selecting Key Performance Indicators Global Integration of Graduate Programmes IV (GIGP) International Conference, pp. 68-75
2. Hellenic Quality Assurance Agency (http://hqaa.gr)
3. H. Small and C. B. Griffith, "The Structure of Scientific Literatures I: Identifying and Graphing Specialties," Science Studies, vol. 4, pp. 17- 40, 1974
4. Katy Börner, Andrea Scharnhorst. Visual conceptualizations and models of science Journal of Informetrics, Volume 3, Issue 3, July 2009, Pages 161-172.
5. Chen, C. (2006) CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology, 57(3), 359-377.
6. V. Batagelj, "Efficient Algorithms for Citation Network Analysis," in arXiv:cs/0309023v1, 2003.
7. Anastasios Tsolakidis, Ioannis Chalaris and George Miaoulis : Interactive Visualization in Knowledge Discovery for Academic Evaluation and Accreditation, Lectures Notes in Computer Science, 2009, Vol. 5914, Knowledge Science, Engineering and Management, Pages 168-179.
8. Marko A. Rodriguez, Alberto Pepe : On the relationship between the structural and socioacademic communities of a coauthorship network. Journal of Informetrics Volume 2, Issue 3, Pages 195-201
9. Terttu Luukkonen, Maria Nedeva, Towards understanding integration in research and research policy, Research Policy, 39(2010), No. 5, June 2010, 674-686.
10. Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442
11. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte Etienne Lefebvre - Fast unfolding of communities in large networks Journal of Statistical Mechanics: Theory and Experiment, 1742-5468, P10008 (12 pp.), 2008
12. Xiaodi Huang, Peter Eades, Wei Lai. A Framework of Filtering, Clustering and Dynamic Layout Graphs for Visualization, ACSC '05 Proceedings of the Twenty-eighth Australasian conference on Computer Science - Volume 38
13. **Gephi :** The Open Graph Viz Platform (http://gephi.org/)