Co-authorship networks in academic research communities: the role of network strength

Tsolakidis Anastasios[1,2], Cleo Sgouropoulou[1], Effie Papageorgiou[3], Olivier Terraz[2], Georgios Miaoulis[1,2]

[1]Department of Informatics
Technological Educational Institute of Athens
Ag.Spyridonos st., 122 10 Egaleo, GREECE
Email: atsolakid@teiath.gr, csgouro@teiath.gr, gmiaoul@teiath.gr

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

[3]General Department of Mathematics
Technological Educational Institute of Athens
Ag.Spyridonos st., 122 10 Egaleo, GREECE
Email: efipapag@teiath.gr

***Abstract*—Educational institutions increasingly need to assess and enhance their activities, in order to improve research effectiveness and measure future capability as well as past performance. Research effectiveness is, among others, an important indicator of the quality assurance process in Higher Education. This study examines the research collaboration structures within a Higher Education academic unit, over a four-year period, with the aim of supporting institutional evaluation and reaching a deeper understanding of the process by which academics share research activities. Our approach focuses on the analysis and visualization of the relationships that exist among authors as well as of the researchers' activity and participativeness in intra and inter-institutional research groups.**

Keywords- research indicators; visualization; co-authoring; knowledge discovery; centralities; graph metrics

I. INTRODUCTION

Research within Higher Education has to deal, on one hand, with the increasing and expanding scientific areas and, on the other hand, with the diminishing resources. As a consequence a clear view and recording of the faculty members' research activity is essential for the creation of institutional research profiles that enable the promotion of collaboration arrangements as an effective way to enhance scientific performance and increase the quantity and quality of research outcomes (e.g. articles, patents, etc.)

In our study we examine the academic research collaboration structure, using a co-authoring network, within the Department of Informatics of the Technological Educational Institute of Athens, in order to develop a deeper understanding of the process by which faculty members communicate with each other. Through our analysis we can visualize such a network using graph representation, presenting academic scientific collaborations under the prism of creativity and scientific progress [1]. In addition, we estimate the network strength [2] which reflects the probability for innovation, as one of the main factors for economic growth [18]. Our approach has, so far, been based on the analysis of conference and journal articles as primary research outputs.

On this basis, our study analyses the research activity and collaboration in a depth analysis of a small network of 233 conference and journal articles. We seek to unveil the implied research communities and to find the relationships between the scientific impact and the researchers' position into the collaboration network [1]. Therefore, the research objectives of this study have been:

- to identify established research communities,
- to analyze and map collaborations among the academic researchers, and
- to demonstrate the identification of key researchers by characterizing them as "research hubs" [17].

Research on related work reveals that there is not a unique, optimal way for the representation of related data so as to support responses to the above criteria. Our approach provides a mechanism that enables the analysis and evaluation of academic research collaborations using the centrality measure at the co-authorship network in combination with community detection algorithms to generate the representation of existing research communities.

This paper is organised as follows: Section 2 introduces related work categorised on the basis of applied methodologies. Section 3 presents the working dataset while section 4 focuses on the methods applied in this study discussing the findings and their impact. Finally, sections 5 and 6 conclude this study and indicate future work potential.

II. LITERATURE REVIEW

Co-authorship analysis is a useful metric for exploring collaboration patterns in a Higher Education Institute. It is the most common indicator assuming that co-authorship indicates a level of scientific collaboration [6].

Network analysis uses mathematical models and graph theory to analyze community graphs, e.g. centrality, distance, diameter, and cluster coefficient [7]-[9]. The referenced studies regard either the co-authorship network features or the individual author rankings within the different domains. Bibliometric studies [10] regarding co-authorship have focused mainly on the effects of collaboration to the scientific progress, based on authors as units of analysis. Other studies focus on social network analysis / network science [12], [11], qualitative methods of observation and interviews [13], [14] as well as surveys [15], [16].

Therefore, the evaluation of research activity within a Higher Education academic unit requires measurement along many dimensions and, in many cases, the design and utilization of multidimensional indicators. In our case, emphasis is clearly on researchers' collaborations and the usage of network analysis metrics in order to form an accurate view of the related activity.

III. DATA COLLECTION AND ANALYSIS

The data used for this study have been obtained through our prototype software system which supports the overall process using data visualization [3]. Research articles and their citations have been extracted from the Quality Assurance Unit (QAU) of the Institute, Web of Science, Scopus and Google scholar, for the time period between 2006 and 2009.

The Department of Informatics comprises 3 sectors:

1. Sector of computer programming
2. Sector of information systems and applications
3. Sector of computer systems and networks

No. of papers	251
No. of Authors	25
No. of Citations	276
No. of Projects	51
Sectors	1. Sector of computer programming
	2. Sector of information systems and applications
	3. Sector of computer systems and networks
Faculty post	1. Professor
	2. Assistant Professor
	3. Associate Professor
	4. Lecturer

TABLE 1: EVALUATION DATA FOR THE PERIOD 2006-2009

As recorded in Table 1, 251 articles were analysed: 144 have an authoring team including only one faculty member of the department, while the rest 107 articles comprise 63 dual-authorship papers, 43 triple-authorship papers and 1 four-author paper. For our experiment we constructed a binary 25*25 matrix where each author constitutes a node. Hence, the resulting co-authorship network contains 25 nodes (authors) connected by 107 collaboration ties, with an Average Degree of 4.28, meaning that the authors have few relationships with each other.

IV. ANALYSIS AND RESULTS

A. Research Communities Detection

In order to detect the communities that may exist in our network (co-authorship graph) we use the Louvain method. A community is a cluster of nodes which has strong connections to each other. The Louvain method is a Multi-Level Aggregation Method for optimizing modularity [4]. The method consists of two phases. Initially, it looks for "small" communities by optimizing modularity in the network and then it builds a new network, the nodes of which represent the communities. These steps are repeated iteratively until a maximum of modularity is attained.

The modularity of a partition is a scalar value [-1, 1] that measures the density of links inside communities as compared to links between communities. In the case of our co-authorship network, having weights on the links such as the number of collaborations between and among the authors, the modularity is defined as:

$$Q = \frac{1}{2m}\sum_{i,j}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j),$$

where A_{ij} represents the weight of the edge between i and j, k_i is the sum of the weights of the edges attached to vertex i,

$$k_i = \sum_j A_{ij}$$

c_i is the community to which vertex i is assigned,
$\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and

$$m = \frac{1}{2}\sum_{ij} A_{ij}.$$

Figure 1 depicts the co-authorship network using the Louvain community detection algorithm which deconstructs the network into 6 structural communities. The algorithm assigns a membership value to each of these communities (nodes). This value identifies the degree of collaborations, indicated by a unique color. The diameter of the nodes represents the number of publications, where authors with more publications have larger diameters. In Table 2, we can see the number of authors belonging to these communities and the number of publications.

Sum of faculty members	Modularity Class	Number of Publications
10	6	95
5	5	45
2	4	20
3	3	45
1	2	21
3	1	11

TABLE 2: RESEARCH COMMUNITIES

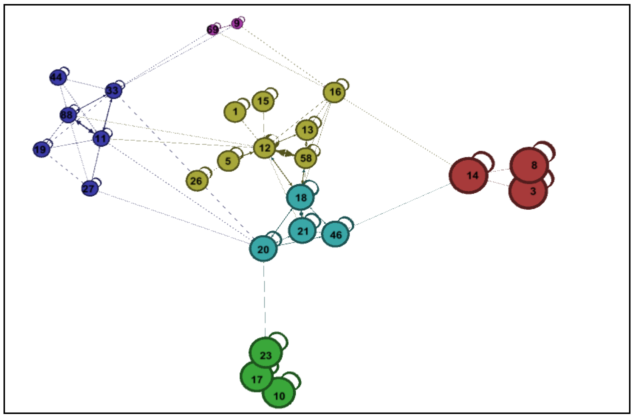

FIGURE 1: CO-AUTHORING NETWORK USING LOUVAIN METHOD

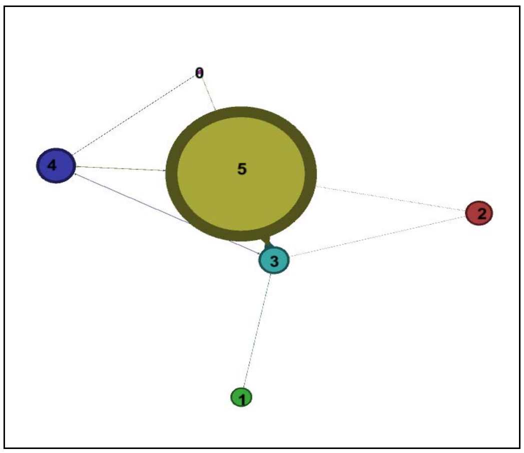

FIGURE 2: GROUP OF COMMUNITIES

As we may see in Figure 2, there is one big, two smaller and three minor communities sharing only a few publications. In addition, there are different types of member relationships within and across communities. More specifically, there may be strong or weak intra-community ties, and direct or indirect inter-community connections.

B. *Network Strength*

In order to define network strength we combine several measures originating from network analysis, such as clustering coefficient [5], graph density and average distance among the nodes in a graph.

1) *Clustering Coefficient*

Clustering coefficient [5] is a measure of the 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 the neighbourhood is fully connected, otherwise there are no connections in the neighbourhood.

The density of clique-like triangles is measured by calculating the clustering coefficient of the network.

$$C_i = \frac{\text{number of triangles connected to node}_i}{\text{number of triaples centered on node}_i}.$$

In the co–authorship network of our case the average clustering coefficient is 0.42 and the total number of triangles is 69. The fact that the average is 0.42 implies that the network is less cliquish. Being very close to 0.5 the measure does not allow us to reach an accurate decision with regard to cliques.

2) *Density*

The density metric in a graph measures the number of edges close to the maximal number of edges. The range of the values is 0 as minimum and 1 as maximum, indicating a complete graph. The density of our network is 0.17 implying a sparse network.

For undirected simple graphs, the graph density is defined as:

$$D = \frac{2|E|}{|V|(|V|-1)},$$

where E represents the edges and V the vertices of the network.

3) *Distance*

The average distance is calculating the shortest path between two pair of nodes. This metric provides a measure of the connectivity among the authors and their ability to collaborate with each other. From the co-author network we observe that the average distance is 4.56. On the grounds that the size of the network is small we argue that distance values are too high. So, in a network of 25 authors an author will need an average of 4.56 steps in order to transfer information to another author. This means that there is a problem in information transfer which is a crucial factor for innovation.

Network strength		**Result**
Avg.Clustering coeffiecient	0.42	Mean
Density	0.17	Low
Avg.Distance	4.56	Low

TABLE 3: ANALYSIS OF NETWORK STRENGTH

C. *Network Analysis*

One of the main issues for designing institutial research policy is to identify the most important researchers among the faculty members who could be the key to enhancing scientific performance and increase the research outcomes (e.g. papers, patents, etc.). On this basis, we use the degrees of Centrality, Betweeness and Closeness in the co-authorship network in order to identify

1. The most active researchers (producing the most research outputs)
2. The researchers with the most collaborations
3. The researchers who act as "research hubs"

1) *Centrality Degree*

Centrality Degree measures the number of lines incident to a node. Authors with high degree centrality are those who have the most collaborations. Using this measure we could identify the most active researchers. In our case study the participating researchers' values range between 0 to 22. For example, the researcher with id=12 has the highest value in the network, having established 22 collaborations with the others.

The Centrality Degree is defined as follows:

$$C_D(n_i) = d(n_i),$$

where $d(n_i)$ is the degree of n_i.

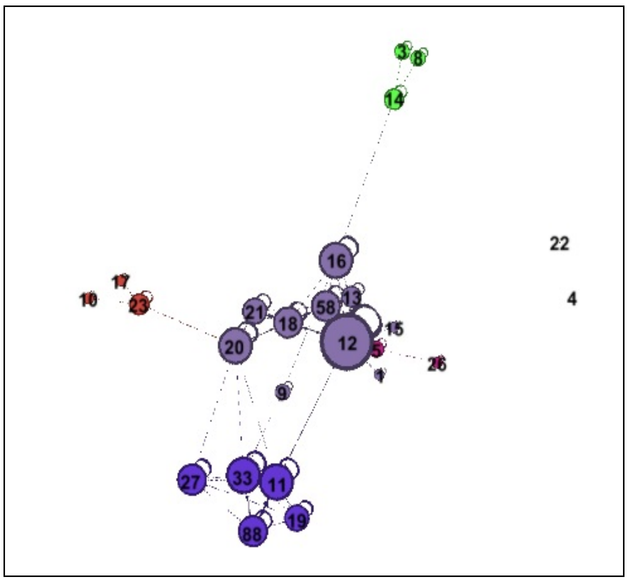

FIGURE 3: CO-AUTHORING NETWORK USING CENTRALITY DEGREE

2) *Betweeness Degree*

Betweenness Degree measures the ability of a node to connect nodes that do not have have any direct connection (edge). These nodes are called hubs, because they have the capacity to transfer information from one researcher to another. The values are ranging from 0.38 to 0. In Figure 4 we can observe that the author with id=12 has the higher value of betwenness.

The betweenness centrality of a node u is given by the expression:

$$g(u) = \sum_{s \neq u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}},$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(u)$ is the number of those paths that pass through u.

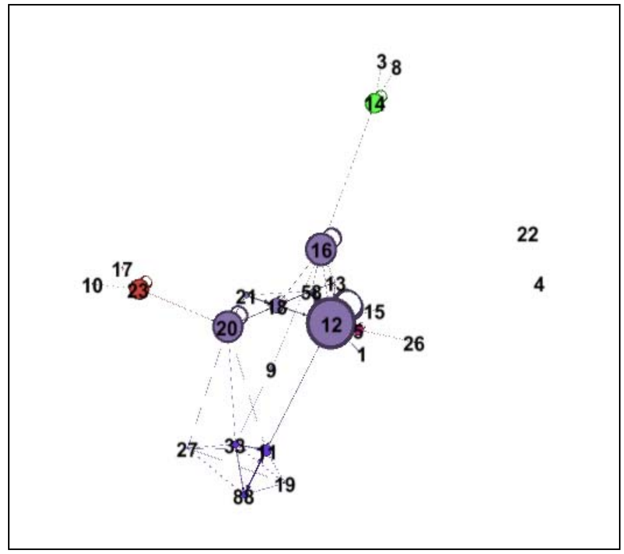

FIGURE 4: CO-AUTHORING NETWORK USING BETWEENESS DEGREE

3) *Closeness Degree*

Closeness centralization is based on the total distance between one node and all other nodes. An author is considered with high closeness centrality if he has many, short connections to other authors in the network [9]. For example authors 18 and 12 have the highest closeness values, meaning that they tend to collaborate easier as they have the shortest paths to the other nodes.

The Closeness Centrality $C_c(n_i)$ is given by the expression:

$$C_C(n_i) = \sum_{i=1}^{n} \frac{1}{d(n_i, n_j)},$$

where (n_i, n_j) is the distance between two vertices in the network.

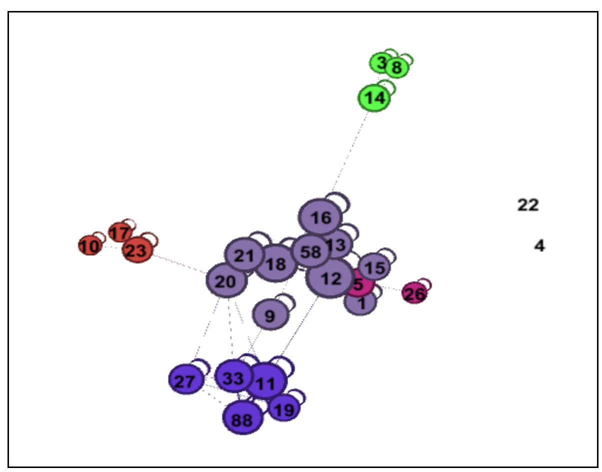

FIGURE 5: CO-AUTHORING NETWORK USING CLOSENESS DEGREE

Criteria	Network Degrees	Ranking
The most active researchers, with the most research outputs	Centrality Degree	Author with id=12
The researcher with the most collaborations among the others	Closeness centralization	Author with id=12
The researchers who has the role of "research hubs"	Betweeness Degree	Author with id=12

TABLE 4: NETWORK ANALYSIS

V. CONCLUSIONS – RESULTS

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. Graph analysis can be effectively used on analysing research outputs in universities for the identification of research communities, the most active researchers, the "research hubs" and also the strength of the co-authorship network which reflects the ability for scientific progress. The analysis of a department's research articles provides a detailed insight to the relationships among its faculty members. The structural analysis of the co-authoring network illustrates the research relationship of the scientists, using different graph metrics. In our case, findings indicate that the relationship and communication among the faculty members is weak. In the department of 25 academic researchers we detect 6 different communities having significant differences in the number of research outputs. In addition, we observe that the top four researchers in the authors' ranking (centralities) are the ones with the most publications, pointing out that the number of research outputs clearly depends on the established collaborations.

VI. FUTURE WORK

There are currently several limitations in our research that should be addressed in the future work. The results are based on data limited to research output. The analysis presents the research papers and journal articles within the faculty members of the department. The next step of our research is to enrich the data with topics of research in order to evaluate the development of scientific areas in comparison to the publications. Potentially, it is possible and necessary to add semantics and thus construct a systematic model for evaluation.

REFERENCES

[1] Ling Cao , Pauleen, D. ; Wang, W.Y.C. ; Whitworth, B. Structural Analysis in the Collaborative Research Network -- The Empirices of Chinese Meteorology Researchers, Computational Sciences and Optimization (CSO), 2011

[2] Eisingerich, A.B., Bell, S.J., Tracey, P. How can clusters sustain performance? The role of network strength, network openness, and environmental uncertainty (2010) Research Policy, 39 (2), pp. 239-253.

[3] Tsolakidis,A.; Sgouropoulou, C.; Xydas, I.; Terraz, O.; Miaoulis, G.; (2011) ' Academic Research Policy-making and Evaluation using Graph Visualisation', 15th Panhellenic Conference on Informatics (PCI), p28 - 32

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

[5] Matthieu Latapy, Main-memory Triangle Computations for Very Large (Sparse (Power-Law)) Graphs, in Theoretical Computer Science (TCS) 407 (1-3), pages 458-473, 2008

[6] Newman M. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the USA (PNAS) 2004;101(suppl. 1):5200–5205.

[7] Grossman JW. The evolution of the mathematical research collaboration graph. Congressus Numerantium. 2002;158:202–212.

[8] Barabási AL. Linked: How Everything Is Connected to Everything Else and What it Means for Business, Science, and Everyday Life. 2002 2002. ISBN 0-452-28439-2.

[9] Milojevic S. Modes of collaboration in modern science: Beyond power laws and preferential attachment. Journal of the American Society for Information Science and Technology.2010;61(7):1410–1423.

[10] Glanzel W, Schubert A. Analysing scientific networks through co-authorship. In: Moed HF, Glanzel W, Schmoch U, editors. Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems. 2004. pp. 257–276.

[11] Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. Research Policy, 34(10), 1608-1618.

[12] Barabási, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. Physica A, 311, 590-614.

[13] Shrum, W., Genuth, J., & Chompalov, I. (2007). Structures of Scientific Collaboration. Cambridge: MIT Press

[14] Hara, N., Solomon, P., Kim, S.-L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the American Society for Information Science and Technology, 54(10), 952-965

[15] Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. Journal of the American Society for Information Science and Technology, 57(13), 1758-1770

[16] Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. Social Studies of Science, 35(5), 673-702

[17] Larsen, K. 2008. Knowledge network hubs and measures of research impact, science structure, and publication output in nanostructures solar cell research. Scientometrics, Vol. 74, no.1: 123-142

[18] European Innovation Scoreboard 2008. Comparative Analysis of Innovation Performance

Appendix A.

Construct	Measures
Network strength	Avg.Clustering coeffiecient Density Avg.Distance
Network Analysis	Centrality Degree Closeness centralization Betweeness Degree

Appendix B.

Id	Modularity Class	Closeness Centrality	Number of Publications	Degree	Clustering Coefficient	Betweenness Centrality
1	6	0,35	4,00	4,00	0,14	0.0
9	6	0,40	6,00	6,00	0,53	0,02
12	6	0,53	63,00	22,00	0.6	0,38
13	6	0,39	11,00	8,00	0.8	0.0
15	6	0,35	10,00	4,00	0,66	0.0
16	6	0,47	10,00	14,00	0.5	0,24
18	6	0,48	40,00	12,00	1.0	0,10
20	6	0,44	13,00	14,00	1.0	0,23
21	6	0,43	16,00	10,00	0.0	0,03
58	6	0,44	33,00	12,00	0.0	0,01
11	5	0,47	22,00	14,00	0,66	0,08
19	5	0,34	4,00	10,00	0,33	0.0
27	5	0,38	2,00	12,00	1.0	0,01
33	5	0,42	11,00	14,00	1.0	0,04
88	5	0,44	19,00	12,00	0.0	0,04
5	4	0,36	20,00	6,00	0,19	0,07
26	4	0,27	1,00	4,00	0.0	0.0
10	3	0,25	16,00	4,00	0,66	0.0
17	3	0,25	35,00	4,00	0,33	0.0
23	3	0,33	26,00	8,00	1.0	0,14
22	2	0,00	0,00	0,00	0.0	0.0
3	1	0,26	11,00	6,00	0,42	0.0
8	1	0,26	2,00	6,00	0,38	0.0
14	1	0,34	19,00	8,00	0.6	0,14
4	0	0,00	0,00	0,00	0,23	0.0

TABLE 5: CO-AUTHORSHIP ANALYSIS