Program

External collaboration patterns of research institutions using shared publications in the Web of Science

Sergio Luis Toral, Nik Bessis, María del Rocío Martínez-Torres,

Article information:

Received 27 May 2012
Revised 9 September 2012
Accepted 11 November 2012

External collaboration patterns of research institutions using shared publications in the Web of Science

Sergio Luis Toral
Departamento de Ingeniería Electrónica, Universidad de Sevilla, Sevilla, Spain
Nik Bessis
School of Computing and Maths, University of Derby, Derby, UK, and
María del Rocío Martínez-Torres
Departamento de Administración de Empresas y Comercialización e Investigación de Mercados (Marketing), Universidad de Sevilla, Sevilla, Spain

Abstract

Purpose – During recent decades, research institutions have increased collaboration with other institutions since it is recognized as a good practice that improves their performance. However, they do not usually consider external collaborations as a strategic issue despite their benefits. The purpose of this paper consists of identifying different patterns of collaboration and internationalization of universities, with the aim of helping managers and policy makers to take decisions related to their national research policies.

Design/methodology/approach – Co-authorship analysis has been used in conjunction with social network analysis to model inter-institutional collaborations as networks, extracting these collaborations from the Web of Science database. Using several structural properties of the extracted networks and applying a statistical treatment, the main profiles of collaborations and internationalization have been obtained.

Findings – Obtained results distinguish three patterns of collaborations according to the intensity and scope of collaborations. The statistical treatment also provides a segmentation of universities according to their collaboration profiles. Finally, universities are represented in bi-dimensional maps using external collaborations as a measure of similarity.

Research limitations/implications – Although this study is restricted to English universities, it could probably be extended at least to other countries in the European Union or even other developed countries.

Practical implications – Research and institutions productivity are usually linked to the amount of received funding. The use of indicators related to internationalization of institutions can help to avoid a bias in favour of research quantity rather than quality, and towards a short-term performance rather than a long-term research capacity.

Originality/value – As a difference to previous works, this paper analyses networks of collaboration from the viewpoint of institutions. More specifically, the combination of social network analysis and factor analysis is used to identify patterns of collaboration among institutions. A longitudinal study is also included to demonstrate that the obtained categorization of universities is maintained over time.

Keywords Knowledge networks, External collaborations, Internationalization, Social network analysis, Bibliometrics, Collaborative research networks, Universities, United Kingdom

Paper type Research paper

Program: electronic library and information systems
Vol. 47 No. 2, 2013
pp. 170-187

0033-0337
DOI 10.1108/00330331311313753

1. Introduction

Co-operation between different scientific disciplines, different organisational units, and external actors seem to be a common and increasing phenomenon of academic reality (Guimera *et al.*, 2005). Obviously, collaboration has always been part of academic life, but the context of an increasingly globalised research environment has encouraged academic institutions to strengthen their external and international dimension. Both of them are considered essential to remain competitive and to drive economic growth. Several reasons can explain this trend. Research collaboration is important in order to meet the big global challenges confronting science. In fact, an increasing number of topics require today an interdisciplinary treatment, being necessary the participation of department or organisation belonging to different fields of knowledge (Bessis, 2009; Bordons *et al.*, 1999; Knowles, 2010). From the researchers point-of-view, group collaborations show that people who already have written a paper previously together are much more likely to succeed in future collaborations, as they have already paid the start-up costs of getting to know each other's languages, approaches and methodologies (Cummings and Kiesler, 2008). Furthermore, attracting and retaining links with the best scientific talents ensures that researchers and institutions stay at the centre of global innovation networks.

In general, all the actors involved in the academic world, including researchers, managers, politicians and policy makers, are broadly in agreement about the benefits of reinforcing external collaborations. The main benefits for higher education institutions derive from being visible and attracting reputed researchers as well as potential research students from all over the world. Researchers also consider more exciting in working with people and groups that have different skills and viewpoints. Politicians and policy makers are usually more worried about the competitiveness and sustainability of the domestic research system and the domestic economic growth.

The most extended technique for analysing collaborations is co-authorship analysis (Liu *et al.*, 2005). Much of the previous work in this area has used co-authorship analysis to assess collaboration among researchers or the structure of scientific collaborations (Farkas *et al.*, 2002; Rodriguez and Pepe, 2008). In this paper, we propose the use of co-authorship analysis to assess the collaboration among research institutions in a certain geographical environment using social network analysis techniques. As a case study, academic institutions of England are considered, and thus, the extracted results have been analysed using social network analysis techniques. The remainder of this paper is presented as follows: section 2 details previous research in the field of co-authorship analysis and social network analysis. Section 3 describes the methodology, including the set of considered data, retrieval of information and the analysis of this information. Obtained results are presented in section 4 while their discussion and implications are included in section 5. Finally, the conclusions can be found in section 6.

2. Research framework

Co-authorship networks represent a class of social networks typically used to determine the structure of scientific collaborations and the status of individual researchers. These networks are usually analysed using bibliometric methods. Although they are somewhat similar to the much studied citation networks (Garfield, 1979), co-authorship implies a much stronger link than citation, which can occur without the authors knowing each other (Wang *et al.*, 2011).

Regarding co-authorship, several previous works agree that collaboration of individual scientists and that of institutions or of even higher levels of aggregation have to be clearly distinguished (Laudel, 2002; Katz and Martin, 1997). In this context, institutional collaboration can, in turn, be studied in two important aspects: collaboration between different research institutions disregarding their organisational type, and collaboration between different sectors such as university, industry, and government (Glänzel and Schubert, 2004). This work is focused on the first one, sometimes called inter-institutional collaboration. An analysis of scientific collaboration from an inter-institutional point-of-view can contribute to define the ranking position of academic institutions in a specific geographical area or in a specific knowledge domain (Olmeda-Gómez *et al.*, 2009). Most studies of inter-institutional collaboration are restricted to national or regional analyses (Hicks and Katz, 1997; Gómez *et al.*, 1995). For instance, a study of domestic inter-institutional collaboration in Canada, Australia, and the UK has concluded that research co-operation decreases exponentially with the distance separating the collaborative partners (Katz, 1994). As a difference, international collaboration is not only determined by distance, but also by other factors such as the country size and political and economic reasons, as well as certain aspects of mobility and migration at the individual level (Glänzel and Schubert, 2004). Other authors consider that innovations in information and communication technologies have removed some of the barriers to collaboration causing a proximity effect between researchers even if they are geographically far away (Katz and Martin, 1997; Li and Robertson, 2011). Besides, there are also strong influences of historical, cultural and linguistic proximities on co-operation patterns at the national and international level (Beaver, 2001).

In general, scientific collaboration is accepted as a basically positive phenomenon and is unanimously recognised as exerting a significant influence on the performance of individual researchers and institutions, in terms of both effectiveness and efficiency (Abramo *et al.*, 2009). Several studies conclude that collaborations contribute to scientific productivity (Lee and Bozeman, 2005). In this study, authors attempted to evaluate the degree to which collaboration among scientists influences scientific productivity, as measured in terms of publications. The results showed that the number of collaborating researchers is the strongest predictor of productivity and that the positive correlation between collaboration and productivity is adequately robust. More specifically, international collaboration has been commented for producing real and remarkable results in the scientific performance of research groups (Martin-Sempere *et al.*, 2002; Barjak and Robinson, 2007).

The application of social network analysis to a co-authorship networks scenario has become increasingly common during the last decade (Olmeda-Gómez *et al.*, 2009). They share with other social networks global topological properties such as small world-property, long-tail degree distribution and a scaling law for the clustering coefficient (Velden *et al.*, 2010). In a similar vein, several previous works have analysed the relationships among researchers in specific geographical areas (Lariviere *et al.*, 2006) or specialities (Hou *et al.*, 2008). In particular, they model researchers as nodes of the social networks and extract some conclusions from the links of collaborations. As a difference, some other works model institutions as nodes, visualizing inter-university and international collaboration networks. For instance, a microanalysis of inter-institutional co-authorship networks comprising universities, government and private companies located in Madrid, Spain, is proposed in Olmeda-Gómez *et al.* (2008).

3. Hypotheses

In this work, academic institutions are also modelled as the nodes of the co-authorship network in a specific geographical area (England). However, three partitions of the global network are going to be considered: England universities, other England research institutions and external (outside of England) research institutions. The purpose of this research consists of identifying similarities and dissimilarities in the collaboration policies of England institutions, and their impact on the basis of their overall performance.

Previous studies highlight the international nature of a small group of successful institutions that represent the leading edge of higher education's embrace of the forces of globalisation (Mohrman *et al.*, 2008). These universities and research institutions are characterised by an intensity of research that far exceeds past experience. They are engaged in worldwide competition for students, faculty, staff, and funding, and look beyond the boundaries of the countries in which they are located to define their scope as trans-national in nature. This activity is also encouraged by governmental research investment policies, which are increasingly oriented to facilitate partnerships between research universities and corporate entities (Salamon, 2002; Tierney, 2006). In this line, Baba *et al.* (2009) also propose a classification of research organisations in which they distinguish those organisations with a publications record above the average. Therefore, we posit:

H1. Successful research institutions are characterized by their intensity of external research collaborations and their scope of collaborations.

With regard to the benefits of collaboration, the study developed by Melin (2000) over 195 university professors revealed that scientists collaborate for pragmatic reasons, being one of the most relevant the higher scientific quality results. It seems reasonable to assume that in international collaborations, precisely because of the differences between partners, the expected results would be greater (Abramo *et al.*, 2011). Studies of group creativity confirm that it is diversity rather than conformity that leads to more innovative and higher quality results (De Dreu and West, 2001). Researchers from different nations who collaborate together have more probability of learning new (to them) notions, techniques and methodologies from one another, and thus of increasing their personal knowledge assets (Burt, 1992). Therefore we posit:

H2. Universities with the highest collaboration rate are also the ones that exhibit better quality in their collaborations.

Geographical proximity is frequently claimed to be beneficial for successful collaboration and knowledge exchange (Anselin *et al.*, 2000). This is most often explained by the importance of face-to-face contacts for the exchange of tacit knowledge. In this line, the study of Olmeda-Gómez *et al.* (2009) about Spanish universities concluded that Spanish inter-university collaboration patterns appear to be influenced by both geographic proximity and administrative and political affiliation. In this paper, we extend this idea to the patterns of collaboration of successful research institutions, both at national and international level. This group of most relevant institutions pursuits not only a high collaboration intensity but also a high quality of collaboration. Therefore, they tend to collaborate with the same

relevant national research institutions and, through this interaction, with similar international research institutions. As a result, we posit:

H3. Patterns of collaboration of successful research institutions are similar for geographically closer institutions.

In this paper, source of data are joint publications among England universities and other research institutions all over the world. This information is analysed by modelling collaboration among institutions as a social network and extracting several topological features. To obtain the main profiles of external collaboration policies in England universities, a factor analysis has been applied to the data set obtained from the topological characteristics of the derived network. Factor analysis provides not only a tool for the identification of latent dimensions but also a categorization of the data sample according to these dimensions. In the case of our work, this means a categorization of England universities. Visually interpretable results to help managers and policy makers to take decisions about the national research system will be also provided.

4. Methodology and data

A social network can be represented as a graph $G = (V, E)$ where V denotes a finite set of nodes and E denotes a finite set of edges such that $E \subseteq V \times V$. Some network analysis methods are easier to understand when graphs are conceptualised as matrices (Stefanonea and Gay, 2008; Martínez-Torres *et al.*, 2010; Toral *et al.*, 2009a):

$$M = (m_{i,j})_{n^*n} \text{ where } n = |V| \; m_{i,j} = \begin{cases} 1 \text{ if } (v_i, v_j) \in; E \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

In the context of co-authorship, collaborations can be modelled using this kind of networks. As the purpose of this work is about analysing patterns of collaborations among institutions, nodes of the network represent those institutions or organizations in which authors are affiliated, and edges are set whenever institutions share a common work.

Data has been extracted from databases contained on the Web of Science. More specifically, records corresponding to England research institutions during the year 2010 were downloaded. The resulting network is shown in Figure 1. This network contains 9,344 nodes corresponding to England and foreign research or/and research organisations collaborating through almost 213,000 papers: non-England institutions (6,693), shown as white nodes in Figure 1, England research institutions except universities (2,518), and England universities (133), both of them shown as black nodes in Figure 1.

Using the three extracted partitions, the following networks will be considered: the complete network (CN), which includes the whole data set of 9,344 nodes; the England institutions network (EN), which only considers collaborations among England institutions (excluding foreign institutions); and the England universities plus Foreign institutions network (UFN), which is focused on the international collaborations of England universities.

Several features of these networks can be calculated through social network analysis (SNA) techniques:

- *Density*. It is defined as the number of lines in a simple network, expressed as a proportion of the maximum possible number of lines. The main problem of this

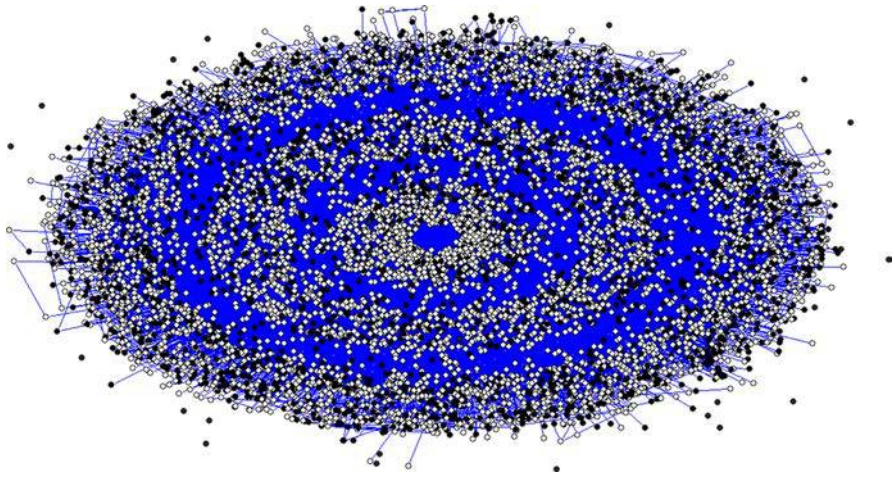


Figure 1.
Collaboration network of English research institutions in 2010

definition is that it does not take into account valued lines higher than 1 and it depends on the network size. A different measure of density is based on the idea of the degree of a node, which is the number of lines incident with it (Toral *et al.*, 2009b). A higher degree of nodes yields a denser network, because nodes entertain more ties, and the average degree is a non-size dependent measure of density.

- *Closeness centralization*. It is an index of centrality based on the concept of distance. The closeness centrality of a node is calculated considering the total distance between one node and all other nodes, where larger distances yield lower closeness centrality scores. The closeness centralization is an index defined for the whole network, and it is calculated as the variation in the closeness centrality of vertices divided by the maximum variation in closeness centrality scores possible in a network of the same size (Toral *et al.*, 2009b).
- *Brokerage roles*. A broker is a middle node in a directed triad (a set of three vertices and the lines among them). Different types of brokerage roles can be distinguished considering mediation between different types of nodes. In the context of this study, brokerage roles among nodes of the three extracted partitions can be considered separately (Toral *et al.*, 2010).
- *Neighbours*. Number of nodes adjacent to each node.
- *Clustering coefficient*. It measures whether first degree neighbour of a particular node interact with each other. Basically, clustering coefficient is a measure of local cohesiveness through the neighbour interactions of a node.
- *Structural holes*. They refer to the extent a node performs a bridging role among a set of nodes that are not directly linked (Nooy *et al.*, 2005).

5. Results

Table I details the set of indicators that have been measured in this study. DegreeCN refers to the degree of each node of the complete collaboration network. Degree value considers multiple lines, that is, multiple collaborations between the same institutions. Consequently, the degree shows not only multiple collaborations with other

Indicator	Description
DegreeCN	Degree of each node (complete network)
ClosenessCN	Closeness centralization (complete network)
ClosenessUEN	Closeness centralization (England research institutions network)
ClosenessUFN	Closeness centralization (universities and foreign institutions network)
CoordCN	Number of brokerage roles among nodes belonging to the same partition (complete network)
ItinUEN	Number of brokerage roles among nodes belonging to the different partitions (England research institutions network)
ItinUFN	Number of brokerage roles among nodes belonging to the different partitions (universities and foreign institutions network)
Nlines1neighCN	Number of lines with adjacent nodes (complete network)
Nlines1neighUEN	Number of lines with adjacent nodes (England research institutions network)
Nlines1neighUFN	Number of lines with adjacent nodes (universities and foreign institutions network)
SHolesCN	Structural holes (complete network)
SHolesUEN	Structural holes (England research institutions network)
SHolesUFN	Structural holes (universities and foreign institutions network)
CC_CN	Clustering coefficient (complete network)
CC_UEN	Clustering coefficient (England research institutions network)
CC_UFN	Clustering coefficient (universities and foreign institutions network)

Table I.
Set of measured indicators

institutions but also repeated collaborations with the same institutions. ClosenessCN is the closeness centrality of a node, which measures multiple collaborations with other institutions but without considering multiple lines. ClosenessUEN and closenessUFN are the closeness centrality of nodes of the England research institutions network and the England universities plus foreign institutions network, respectively. CoordCN is the number of brokerage roles developed by each node but among other nodes belonging to the same partition. As a difference, ItinUEN and ItinUFN refer to the number of brokerage roles developed by each node but among other nodes belonging to the different partitions. In the case of ItinUEN, partitions considered are universities and the rest of England research institutions whole in the case of ItinUFN, partitions considered are universities and foreign research institutions. Nlines1neighCN, Nlines1neighUEN, Nlines1neighUFN represent the number of lines with adjacent nodes for the three considered networks. It is a measure if the intensity of collaboration. Finally, SHolesCN, SHolesUEN, SHolesUFN measures the extent each node behaves as a structural hole in each of the three considered networks, while CC_CN, CC_UEN and CC_UFN are the corresponding clustering coefficients.

The set of indicators shown in Table I have been measured for the list of 133 England universities corresponding to one of the extracted partitions. Those universities with a zero degree value have been excluded, leading to the total of 115 universities.

First of all, a ranking of England universities has been used to test the influence of the listed indicators in the overall performance of universities. For this purpose, the ranking of the Complete University Guide (www.thecompleteuniversityguide.co.uk/league-tables/rankings) has been chosen. In particular, this guide offers a ranking of England universities attending to their average quality of their research. Table II shows the correlation values of this ranking with the set of extracted indicators. Notice

that a negative value means a high correlation with the performance of universities in terms of research, as usually rankings are ordered from lower to higher values.

According to Table II, performance of universities is positively correlated with external collaborations, both in terms of internal and international collaborations. It is also positively correlated with the scope and intensity of these collaborations. On the other hand, performance is not significantly correlated with the role of universities as structural holes and negatively correlated with their clustering coefficient.

To obtain the different patterns of collaborations of England universities, a factor analysis has been applied to the extracted data set. Factor analysis is a way to fit a model to multivariate data, estimating their interdependence. It addresses the problem of analysing the structure of inter-relationships among a number of variables by defining a set of common underlying dimensions, the factors, which are not directly observable, segmenting a sample into relatively homogeneous segments (Rencher, 2002). Because each factor may affect several variables in common, they are known as "common factors". Each variable is assumed to be dependent on a linear combination of the common factors, and the coefficients are known as loadings. Factor analysis can be used for either exploratory or confirmatory purposes: exploratory analyses do not set any a priori constraints on the estimation of factors or the number of factors to be extracted, while confirmatory analysis does (Raban and Rabin, 2009). In our case, we have developed an exploratory analysis as we did not know the number of underlying dimensions. That is to say, a decision must be made about the number of factors to be extracted. There are several criteria for doing this, being the most extensive the eigenvalue and percentage of variance criterion. The eigenvalue criterion considers a number of factors equals to the number of eigenvalues higher than 1. The percentage of variance criterion considers all factors accounting for about 70 per cent of the variance of the original variables (Rencher, 2002; Martínez-Torres and Toral, 2010). Table III details the eigenvalues and the percentage of variance explained for the proposed case study. Both criteria are satisfied for a number of factors equals to three.

Once the number of factors has been determined, the next step is to interpret them according to the factor loadings matrix. The estimated loadings from an unrotated factor analysis fit can usually have a complicated structure. The goal of orthogonal factor rotation is to find a parameterisation in which each variable has only a small number of large loadings, i.e. is affected by a small number of factors. The rotated factor analysis ensures that factors represent unidimensional constructs while preserving the essential properties of the original loadings. The most popular of these techniques is the varimax

	Ranking		Ranking
DegreeCN	−0.577*	Nlines1neighUEN	−0.698*
ClosenessCN	−0.526*	Nlines1neighUFN	−0.606*
ClosenessUEN	−0.434*	SholesCN	0.180
ClosenessUFN	−0.528*	SholesUEN	0.570
CoordCN	−0.646*	SholesUFN	0.189
ItinUEN	−0.493*	CC1_CN	0.438*
ItinUFN	−0.474*	CC1_UEN	0.570*
Nlines1neighCN	−0.606*	CC1_UFN	0.517*

Note: *Correlation is significant at the 0.01 level (two-tailed)

Table II. Correlation of the average quality of English universities' research with the set of extracted indicators

Factor	Eigenvalues Total	Percentage of variance	Cumulative %
1	8.868	55.422	55.422
2	3.753	23.457	78.879
3	1.840	11.499	90.378
4	0.572	3.574	93.952
5	0.335	2.097	96.049
...	...	...	...
...	...	...	...
15	0.000	0.001	100.000
16	5.66E-005	0.000	100.000

Table III.
Total variance explained

rotation, which seeks rotated loadings that maximize the variance of the squared loadings in each column of the factor loading matrix (Rencher, 2002).

Factor loadings with varimax rotation are shown in Table IV. Each row represents the factor loadings of each variable. Moving horizontally from left to right across the five loadings in each row, the highest loading has to be identified. All the variables associated in this way with the same factor are hypothesised to share a common meaning that the analyst should discover.

According to Table IV, several factors can be distinguished; these are discussed next.

- *F1*. Collaboration intensity: this factor characterizes those universities with the highest collaboration ratings, both in terms of internal and international collaboration. The high value of factor loadings for the degree, brokerage roles and number of lines with the first neighbour confirms this pattern of behaviour. This group is composed by 15 universities, most of them occupying the first positions of the ranking used to test their research performance.

	Factor F1	F2	F3
DegreeCN	0.967	0.142	−0.152
ClosenessCN	0.367	0.844	−0.165
CoordCN	0.798	0.318	−0.345
ItinUEN	0.949	0.080	−0.080
ItinUFN	0.947	0.067	−0.054
ClosenessUEN	0.249	0.911	−0.087
ClosenessUFN	0.373	0.842	−0.162
CC1_UEN	−0.335	0.151	0.747
CC1_UFN	−0.196	−0.174	0.919
SHolesCN	−0.059	−0.970	−0.005
SHolesUEN	−0.030	−0.945	0.096
SHolesUFN	−0.066	−0.971	0.009
Nlines1neighCN	0.926	0.191	−0.219
Nlines1neighUEN	0.847	0.318	−0.346
Nlines1neighUFN	0.923	0.193	−0.221
CC1_CN	−0.149	−0.226	0.930

Table IV.
Rotated factor loadings with Varimax rotation

- *F2*. Scope of collaboration: this group is defined by the high value of closeness centrality and the low value of structural holes (negative factor loadings). That means this group exhibit a good connectivity with other institutions but with lower intensity than the previous group. This group is integrated by 55 universities.
- *F3*. Absence of collaboration: this group is defined by the high value of their clustering coefficient. The clustering coefficient depends on the connectivity of neighbours. In this case, its high value can be interpreted as nodes with a low number of neighbours, which is also justified by the low value of factor loadings corresponding to the degree or the brokerage roles. Up to 33 universities are included in this group.

In addition to factor loadings, which have been used for interpreting obtained factors, factor scores used to categorize the original sample of universities, which can be approximated to one of the identified latent factors. Consequently, the original sample of England universities can be categorized in three groups. An analysis of variance (ANOVA) has been applied to the categorization of the original sample in the three groups obtained form factor analysis. The aim of this analysis consists of checking the null hypothesis of equal population means. Table V details the *F* statistic, the ratio of two different estimators of population variance, which appears together with its corresponding critical level or observed significance. The outcome is that the null hypotheses have been rejected in all the cases with a significance value below 0.05. That means the obtained categorization from factor analysis is well defined.

According to this categorisation, there are 14 universities categorised on factor 1, 55 in factor 2 and 33 in factor 3, while 13 universities could not clearly be approximated to any factor.

Obtained results show that there is a small group of top universities which concentrates the majority of collaborations. Actually, they accumulate about 68 per cent of external collaborations of all England universities. In the case of international

	F	Sig.
DegreeCN	61.002	0.000
ClosenessCN	73.331	0.000
CoordCN	78.137	0.000
ItinUEN	38.029	0.000
ItinUFN	31.260	0.000
ClosenessUEN	49.075	0.000
ClosenessUFN	73.163	0.000
CC1_UEN	30.333	0.000
CC1_UFN	82.128	0.000
SHolesCN	26.258	0.000
SHolesUEN	20.928	0.000
SHolesUFN	26.943	0.000
Nlines1neighCN	116.145	0.000
Nlines1neighUEN	142.634	0.000
Nlines1neighUFN	111.995	0.000
CC1_CN	68.789	0.000

Table V. Statistical significance of ANOVA

collaborations, this percentage increases till 84.7 per cent. Figure 2 illustrates the network of England universities, where the area of nodes is proportional to their degree. In fact, degree follows a power law distribution typical of web-based interactions (Raban and Rabin, 2009). This figure visually highlights the dominant group of universities in terms of collaborations with other institutions. The second group of universities is responsible of almost the rest of collaborations. Although they collaborate with a good number of other institutions, the intensity of this collaboration is still far from those of top universities. Finally, the third group does not exhibit any research orientation nor collaboration strategy.

To demonstrate this categorisation of universities is maintained over time, a longitudinal approach has been considered by repeating the same analysis for the year 2006. In this case, the size of the network is of 7,052 nodes, including the same 133 England universities but a total of 2,106 England research institutions and 4,813 non-England institutions. These values clearly highlight that collaborations among institutions are increasing at a high rate year by year. The application of factor analysis has lead to the same three factors obtained for the year 2010, with the same set of associated indicators and a cumulative explained variance of 92.4 per cent. Table VI shows the distribution of universities through the three distinguished factors. It can be observed that, with the exception of slight variations, most of Universities remain in the same categorisation than in the year 2010, as hypothesised in *H1*.

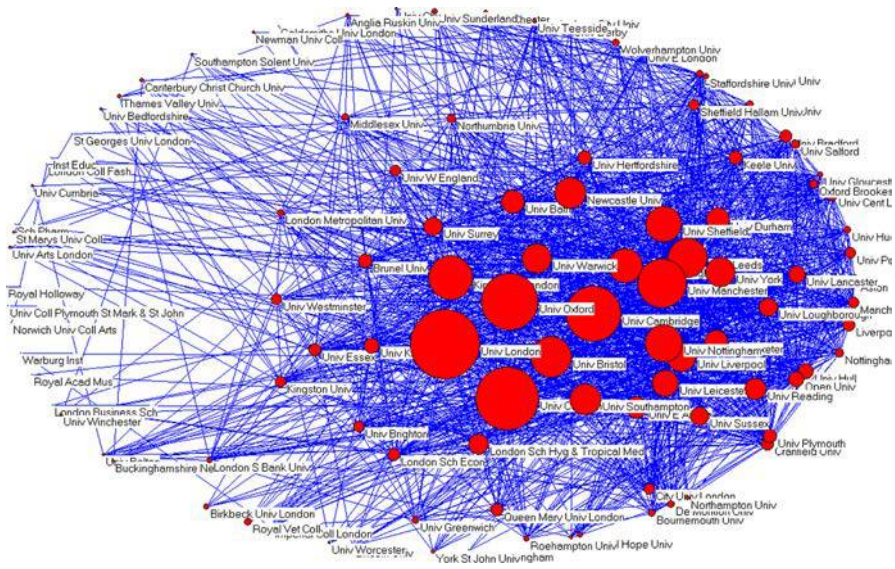


Figure 2.
English universities network

Table VI.
Categorization of English universities for the years 2006 and 2010

	F1	F2	F3
2006	14	55	33
2010	13	57	27

Quality of publications can be incorporated to the extracted network by considering the number of citations associated to joint publications. More specifically, citations are included as multiple lines between nodes. As a result, the degree of nodes includes now the quality of links among research institutions. Data from 2006 have been selected to perform this analysis, as papers need at least three years to accumulate citation data. Table VII compares several statistics related to the degree distribution of nodes corresponding to the collaboration network of England universities during 2006 without considering the quality of publications (first row), 2006 considering the times cited value (second row) and 2010, again without considering the quality of publications (third row). The forth column shows the percentage of collaborations accumulated by research institutions belonging to the F1 group. It can be observed that this percentage remains constant even when the degree value of the network has clearly increased. That means that F1 research institutions accumulate the majority of the highest quality publications, as was posited in *H2*.

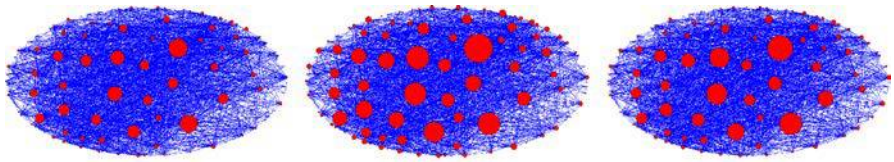
Figure 3 highlights the evolution of the degree of collaboration in the three considered cases. The areas of nodes are proportional to their degree value. It can be observed that the higher average value of the nodes degree is due to the fact that universities with the highest collaboration intensity accumulate more links to the rest of the network.

Although the intensity of collaboration can easily visualized using the size of nodes (Figure 2), the scope of collaboration is more difficult to obtain due to the high density of links among nodes. For this purpose, the 1-neighbour partitions of universities have been extracted. The correlation among these constitutes a measurement of similarities in their external collaboration policies. In order to visually represent these similarities, a multidimensional scaling has been applied (Rencher, 2002). This analysis consists of projecting the similarities on a two-dimensional map, using the data from the correlation matrix as input data.

Figure 4 illustrates the results of applying multidimensional scaling to several cases. These results represent institutions in the form of maps, so those institutions closer in the map are more similar in terms of collaboration policies while those institutions far away means they collaborate with different other institutions. In other words, multidimensional scaling provides a representation of

	Av. value	Median	Maximum value	Percentage of total degree (F1 Institutions)
2006	209.77	26	3,047	66.31
2006TC	471.67	93	6,871	67.75
2010	338.26	44	5,298	71.83

Table VII.
Degree statistics of English collaboration networks



(a) 2006 **(b)** 2006TC **(c)** 2010

Figure 3.
Evolution of the degree of collaboration of English research institutions

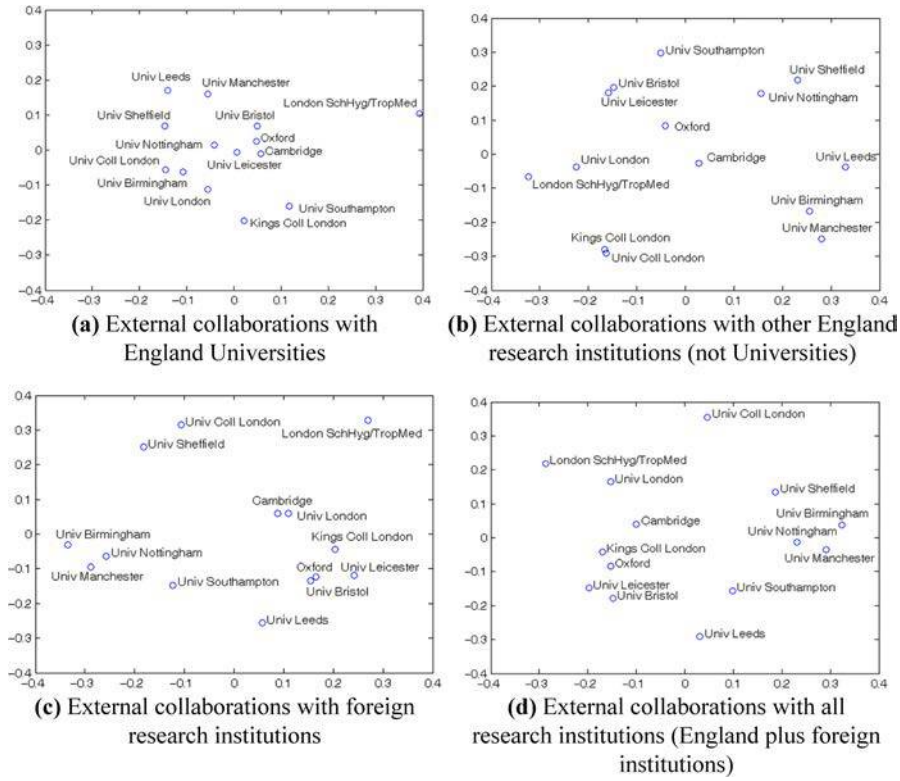


Figure 4.
Multidimensional scaling

the collaboration scope. Figure 4 shows the collaboration scope of 14 universities belonging to F1 considering several subset of data. In particular, Figure 4 (a), (b) and (c) shows the representation for the cases of other England universities, other England research institutions excluding universities and foreign research institutions, respectively. Finally, Figure 4 (d) corresponds to the case of all the previous external institutions.

Obviously, Figure 4 (a) exhibits more similarities as it is more usual that collaboration among universities is located in the same country. In the remaining figures, the distances among universities are higher, which means they exhibit more dissimilarities in their external collaboration policies. In general, it can be appreciated that the geographical distance among universities is related to their external policies. Moreover, analysing the sequence of the four figures, a transversal and longitudinal geographical axis can be distinguished in the set of analysed institutions. The transversal axe correspond to Southern England locations, i.e. Cambridge, Oxford, London, Bristol, while the longitudinal axe correspond to northern locations, Sheffield, Manchester, Nottingham, Birmingham. This trend is more accentuated in Figure 4 (d) where institutions all over the world are considered.

6. Discussion

Today, there is an emerging interest in promoting external collaborations, as it has been proved to have a significant impact in university performance and their ranking position. However, only a minority of universities currently has an international research collaborative strategy although the great majority believes that more strategic management of this activity is needed. Even strategic plans of universities only devote one or two paragraphs to external collaborations and internationalisation policies.

The existence of a small group of top universities is in agreement with previous studies. The analysis of joint publications performed by Calvert and Patel (2003) during the period 1995-2000 leads to a list of top 20 collaborating universities which includes all the universities categorised in group F1 in this work. Moreover, this study also points out that universities with the highest volume of collaborative activity are mainly those with the highest research rankings as hypothesised in *H2*. Several other studies agree with the relationship between collaboration, production and quality of results (De Dreu and West, 2001) due to the potentials linked to the differences between researchers in scientific and cultural background (Abramo *et al.*, 2011). Obtained results in this study specifically highlight intensity and scope as the two parameters able to distinguish more active universities in terms of collaboration. Previous studies (Calvert and Patel, 2003; Mohrman *et al.*, 2008) have been mainly focused in intensity of collaboration by accounting the number of joint collaborations. However, the scope of collaboration is also an important parameter for those middle size universities that do not have the capacity of achieving the level of intensity that other universities do or can do with a much higher number of researchers. Finally, similar patterns of collaboration have been found for geographically closer institutions. Although previous studies also agree with these idea (Olmeda-Gómez *et al.*, 2009), they can also be influenced by the administrative organisations of countries like England or Spain.

From the viewpoint of universities, there is a clear interest in promoting external collaborations, as this activity has a significant impact in their performance and ranking position. In this sense, it is important to know their current situation related to internal and international collaborations. According to our study, is not necessary to maintain a high number of collaborations to be successful. Most universities (factor 2) only maintain a wide variety of partnerships without achieving the high number of collaborations described by factor 1. The most important thing is to choose partnerships aligned with an institution's research and education strategy to guarantee full engagement of research staff (Universities Research Report, 2007). Obviously, benefits of collaboration must be evident for all parties as collaborations require a lot effort in terms of staff time and energy. An institutional support is considered of vital importance for an adequate internationalisation policy (Baskerville *et al.*, 2011). Although there is no standard organisational model, it is recommended a senior university officer with explicit responsibility for international research co-operation (Universities Research Report, 2007). Besides, it is also necessary to provide funds for encouraging external collaborations. Most forms of support are dedicated to international programmes providing small grants to cover the cost of short visits or longer-term exchanges or fellowships. However, this support is intended to help individual researchers develop new relationships with parties in other countries. But they are not usually oriented to support institution-to-institution partnerships. Moreover, funds usually are prioritised to support new relationships rather than to fund more extensive collaboration or to sustain

long-distance relationships in the longer term. In this sense, the proposed methodology can help managers to take decisions based on updated information provided in the form of easily interpretable networks or maps.

From the viewpoint of policy makers, the priority should be to increase the list of top universities to improve the national research system performance. In this sense, external collaboration must be encouraged and linked in some way with interdisciplinary collaboration.

In general, linking indicators with budgetary decisions or promotion decisions affect people in the sense that they react to the implementation of such measures by altering their behaviour. The number of publications, citations received and/or the number of doctoral students supervised are typical indicators used to measure research productivity and decide about future promotions. In turn, researchers are pushed to optimize their productivity levels. One way to achieve this is by splitting research works to the minimum publishable units with the undesirable effect of decreasing the level of completeness in contributions, self-citing or encouraging colleagues to cite each other works. The same can be said for research institutions, whose productivity is also linked to the amount of received funding. However, measures related to internationalisation of institutions are usually not considered by national commission for the evaluation of research activity. But this set of measures offer several advantages. First, researchers and institutions are encouraged to establish collaboration with foreign institutions, as it has been demonstrated that external collaborations usually lead to better quality of works. Second, both the intensity and the scope of the collaboration policy should be considered to avoid manipulations on the intensity of collaboration. The problem of quantity and expansion based funding formulas is that they lead to a growth of collaborations, which in turn, challenges resource scalability both in terms of outputs quality and management. With this in mind, both indicators are considered of equal importance.

In the case of England universities, many HEIs are moving decisively to internationalize their curricula, promote cross-cultural understanding and provide opportunities for the development of foreign-language skills. However, the development of their physical presence and engagements overseas is perhaps the touchstone of their commitment to internationalisation. Some points to be improved are the membership of international networks, the instigation of strategic partnerships and the mobilisation of research teams tasked with the resolution of previously intractable problems. That means that the required internationalisation must be selective, focused, grounded in academic excellence, sustainable for the long term, mutually beneficial and capable of generating complete confidence and trust between the partners.

Although this study is restricted to England universities, it could probably be extended at least to other countries in the European Union or even other developed countries. In fact, prior research agree with the idea of a small group of top universities with a high degree of collaborations, and these studies have considered several different countries. Moreover, globalisation is one of the driving forces of collaboration and this is a phenomenon that has been extended all over the world. Finally, funding in different countries and at the EU level is becoming more and more dependent on the level and intensity of international collaboration, as illustrated by the aim of the EU framework programs to create a "European research area". As a result, all the countries in this area tend to follow similar rules in terms of collaboration.

7. Conclusion

This study has analysed the impact of external collaborations on university performance and has extracted three profiles of universities regarding their external collaborations. Obtained results show a reduced top list of universities, responsible of the majority of collaborations, a second group of universities with lower intensity but wide scope in their external collaboration and a third group of universities with no external collaboration orientation. The scope of collaboration feature has been analysed using multidimensional scaling to distinguish several patterns in the form of maps. Both features intensity and scope of external collaborations are the main features to consider when assessing universities external policies. Obtained results can be easily interpreted and provide valuable information to university managers and policy makers about possible improvements in their strategic policies.

References

Abramo, G., D'Angelo, C.A. and Di Costa, F. (2009), "Research collaboration and productivity: is there correlation?", *Higher Education*, Vol. 57, pp. 155-71.

Abramo, G., D'Angelo, C.A. and Solazzi, M. (2011), "Are researchers that collaborate more at the international level top performers? An investigation on the Italian university system", *Journal of Informetrics*, Vol. 5 No. 1, pp. 204-13.

Anselin, L., Varga, A. and Acs, Z. (2000), "Geographic and sectoral characteristics of academic knowledge externalities", *Papers in Regional Science*, Vol. 79, pp. 435-45.

Baba, Y., Shichijo, N. and Sedita, S.R. (2009), "How do collaborations with universities affect firms' innovative performance? The role of 'Pasteur scientists' in the advanced materials field", *Research Policy*, Vol. 38 No. 5, pp. 756-64.

Barjak, F. and Robinson, S. (2007), "International collaboration, mobility and team diversity in the life sciences: impact on research performance", *Social Geography Discussion*, Vol. 3, pp. 121-57.

Baskerville, S., MacLeod, F. and Saunders, N. (2011), *A Guide to UK Higher Education and Partnerships for Overseas Universities*, UK Higher Education International and Europe Unit, Research Series/9.

Beaver, D. (2001), "Reflections on scientific collaborations (and its study): past, present and prospective", *Scientometrics*, Vol. 52 No. 3, pp. 365-77.

Bessis, N. (2009), *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations*, IGI Publishing, Hershey, PA.

Bordons, M., Zulueta, M.A., Romero, F. and Barrigón, S. (1999), "Measuring interdisciplinary collaboration within a university: the effects of the multidisciplinary research programme", *Scientometrics*, Vol. 46 No. 3, pp. 383-98.

Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge, MA.

Calvert, J. and Patel, P. (2003), "University-industry research collaborations in the UK: bibliometric trends", *Science and Public Policy*, Vol. 30 No. 2, pp. 85-96.

Cummings, J.N. and Kiesler, S. (2008), "Who collaborates successfully?: prior experience reduces collaboration barriers in distributed interdisciplinary research", *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work*.

De Dreu, C.K.W. and West, M.A. (2001), "Minority dissent and team innovation: the importance of participation in decision making", *Journal of Applied Psychology*, Vol. 86 No. 6, pp. 1191-201.

Farkas, I., Derenyi, I., Jeong, H., Neda, Z., Oltvai, Z.N. and Ravasz, E. (2002), "Networks in life: scaling properties and eigenvalue spectra", *Physica A*, Vol. 314, pp. 25-34.

Garfield, E. (1979), *Citation Indexing – Its Theory and Application in Science, Technology, and Humanities*, John Wiley & Sons, New York, NY.

Glänzel, W. and Schubert, A. (2004), "Analysing scientific networks through co-authorship", in Moed, H.F. *et al.* (Eds), *Handbook of Quantitative Science and Technology Research*, Kluwer Academic Publishers, Dordrecht, pp. 257-76.

Gómez, I., Fernández, M.T. and Méndez, A. (1995), "Collaboration patterns of Spanish scientific publications in different research areas and disciplines", in Koenig, M.E.D. and Bookstein, A. (Eds), *Proceedings of the Biennial Conference of the International Society for Scientometrics and Informetrics*, Learned Inf., Medford, NJ, pp. 187-96.

Guimera, R., Uzzi, B., Spiro, J. and Amaral, L. (2005), "Team assembly mechanisms determine collaboration network structure and team performance", *Science*, Vol. 308, pp. 697-702.

Hicks, D. and Katz, J.S. (1997), "The changing shape of British industrial research", STEEP Special Report No. 6, SPRU.

Hou, H., Kretschmer, H. and Liu, Z. (2008), "The structure of scientific collaboration networks in scientometrics", *Scientometrics*, Vol. 75 No. 2, pp. 189-2002.

Katz, J.S. (1994), "Geographical proximity and scientific collaboration", *Scientometrics*, Vol. 31 No. 1, pp. 31-43.

Katz, J.S. and Martin, B.R. (1997), "What is research collaboration?", *Research Policy*, Vol. 26, pp. 1-18.

Knowles, J. (2010), "Collaboration nation: the building of the Welsh repository network", *Program: electronic library and information systems*, Vol. 44 No. 2, pp. 98-108.

Lariviere, V., Gingras, Y. and Archambault, E. (2006), "Comparative analysis of networks of collaboration of Canadian researchers in the natural sciences, social sciences and the humanities", *Scientometrics*, Vol. 68 No. 3, pp. 519-33.

Laudel, G. (2002), "What do we measure by co-authorships?", *Research Evaluation*, Vol. 11 No. 1, pp. 3-15.

Lee, S. and Bozeman, B. (2005), "The impact of research collaboration on scientific productivity", *Social Studies of Science*, Vol. 35 No. 5, pp. 673-702.

Li, J. and Robertson, T. (2011), "Physical space and information space: studies of collaboration in distributed multi-disciplinary medical team meetings", *Behaviour and Information Technology*, Vol. 30 No. 4, pp. 443-54.

Liu, X., Bollen, J., Nelson, M.L. and Van de Sompel, H. (2005), "Co-authorship networks in the digital library research community", *Information Processing and Management*, Vol. 41 No. 6, pp. 1462-80.

Martin-Sempere, M.J., Rey-Rocha, J. and Garzon-Garcia, B. (2002), "The effect of team consolidation on research collaboration and performance of scientists. Case study of Spanish University researchers in Geology", *Scientometrics*, Vol. 55 No. 3, pp. 377-94.

Martínez-Torres, M.R. and Toral, S.L. (2010), "Strategic group identification using evolutionary computation", *Expert Systems with Applications*, Vol. 37 No. 7, pp. 4948-54.

Martínez-Torres, M.R., Toral, S.L. and Barrero, F. (2010), "The role of internet in the development of future software projects", *Internet Research*, Vol. 20 No. 1, pp. 72-86.

Melin, G. (2000), "Pragmatism and self-organization: research collaboration on the individual level", *Research Policy*, Vol. 29 No. 1, pp. 31-40.

Mohrman, K., Ma, W. and Baker, D. (2008), "The research university in transition: the emerging global model", *Higher Education Policy*, Vol. 21, pp. 5-27.

Nooy, W., Mrvar, A. and Batagelj, V. (2005), *Exploratory Network Analysis with Pajek*, Cambridge University Press, New York, NY.

Olmeda-Gómez, C., Perianes-Rodríguez, A., Ovalle-Perandones, M.A. and Moya-Anegón, F. (2008), "Comparative analysis of university-government-enterprise co-authorship networks in three scientific domains in the region of Madrid, 1995-2003", *Information Research*, Vol. 13 No. 3, paper 352, available at: http://InformationR.net/ir/13-3/paper352.html

Olmeda-Gómez, C., Perianes-Rodriguez, A., Ovalle-Perandones, M.A., Guerrero-Bote, V.P. and de Moya Anegón, F. (2009), "Visualization of scientific co-authorship in Spanish universities: From regionalization to internationalization", *Aslib Proceedings: New Information Perspectives*, Vol. 61 No. 1, pp. 83-100.

Raban, D.R. and Rabin, E. (2009), "Statistical inference from power law distributed web-based social interactions", *Internet Research*, Vol. 19 No. 3, pp. 266-78.

Rencher, A.C. (2002), *Methods of Multivariate Analysis*, 2nd ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ.

Rodriguez, M.A. and Pepe, A. (2008), "On the relationship between the structural and socioacademic communities of a coauthorship network", *Journal of Informetrics*, Vol. 2 No. 3, pp. 195-201.

Salamon, L.M. (2002), *The Tools of Government: A Guide to the New Governance*, Oxford University Press, New York, NY.

Stefanonea, M.A. and Gay, G. (2008), "Structural reproduction of social networks in computer-mediated communication forums", *Behaviour and Information Technology*, Vol. 27 No. 2, pp. 97-106.

Tierney, W.G. (Ed.) (2006), *Governance and the Public Good*, SUNY Press, Albany, NY.

Toral, S.L., Martínez-Torres, M.R., Barrero, F. and Cortés, F. (2009a), "An empirical study of the driving forces behind online communities", *Internet Research*, Vol. 19 No. 4, pp. 378-92.

Toral, S.L., Martínez-Torres, M.R. and Barrero, F. (2009b), "Virtual Communities as a resource for the development of OSS projects: the case of Linux ports to embedded processors", *Behavior and Information Technology*, Vol. 28 No. 5, pp. 405-19.

Toral, S.L., Martínez Torres, M.R. and Barrero, F. (2010), "Analysis of virtual communities supporting OSS projects using social network analysis", *Information and Software Technology*, Vol. 52 No. 3, pp. 296-303.

Toral, S.L., Martínez Torres, M.R., Palacios, B. and Barrero, F. (2011), "Web site structure mining using social network analysis", *Internet Research*, Vol. 21 No. 2, pp. 104-23.

Universities Research Report (2007), "International Research Collaboration: Opportunities for the UK Higher Education Sector", Research Report, London.

Velden, T., Haque, A.U. and Lagoze, C. (2010), "A new approach to analyzing patterns of collaboration in co-authorship networks: mesoscopic analysis and interpretation", *Scientometrics*, Vol. 85, pp. 219-42.

Wang, H.-C., Chou, Y.-L. and Guo, J.-L. (2011), "A core journal decision model based on weighted page rank", *Program: electronic library and information systems*, Vol. 45 No. 4, pp. 397-414.

Corresponding author
Sergio Luis Toral can be contacted at: toral@esi.us.es

This article has been cited by:

1. Traci Birch, Enrique Reyes. 2018. Forty years of coastal zone management (1975–2014): Evolving theory, policy and practice as reflected in scientific research publications. *Ocean & Coastal Management* **153**, 1-11. [Crossref]
2. Elisha Chiware, Lara Skelly. 2016. Publishing patterns at the Cape Peninsula University of Technology. *South African Journal of Science* **Volume 112**:Number 1/2. . [Crossref]
3. Mu-Hsuan Huang, Ling-Ling Wu, Yi-Chen Wu. 2015. A study of research collaboration in the pre-web and post-web stages: A coauthorship analysis of the information systems discipline. *Journal of the Association for Information Science and Technology* **66**:4, 778-797. [Crossref]
4. Milorad Pantelija Stevic, Branko Milosavljevic, Branko Rade Perisic. 2015. Enhancing the management of unstructured data in e-learning systems using MongoDB. *Program* **49**:1, 91-114. [Abstract] [Full Text] [PDF]