

# Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature

S. Lozano<sup>\*a</sup>, L. Calzada-Infante<sup>b</sup>, B. Adenso-Díaz<sup>b</sup> and S. García<sup>b</sup>

<sup>a</sup> Department of Industrial Management, University of Seville, Spain, slozano@us.es

<sup>b</sup> Escuela Politécnica de Ingeniería, Universidad de Oviedo, Spain,  
adenso@uniovi.es; calzadalaura@uniovi.es; sgarcia@uniovi.es

\* Corresponding author:

Postal address: Department of Industrial Management  
Escuela Superior de Ingenieros  
Isla de la Cartuja  
41092, Seville, Spain  
E-mail: slozano@us.es  
Phone: +34 954487208

ARTÍCULO PUBLICADO EN LA REVISTA

*Scientometrics* (2019)

doi: <https://doi.org/10.1007/s11192-019-03132-w>

## *Abstract*

A network analysis (NA) of keyword co-occurrences for a broad collection of Data envelopment analysis (DEA) papers in the period 2008-2017 is carried out. The raw keywords have been cleaned up and standardized to consolidate and increase the consistency of the keywords. The resulting network has been characterized using network-level as well as node-level NA measures. Although the size of the network steadily increases with time, the average path length does not, showing its small world character. The disassortativity of the network indicates that the keywords used in a given paper generally include one or more common, frequently-used terms plus other less common terms that refer to the specific context of the research. The evolving nature of the keyword network is highlighted with some DEA keywords staying at the top of the ranking during the whole period and other emerging topics significantly increasing their strength during this period. The community structure of the network, which reflects the field's knowledge structure, is also presented. The identified communities generally include specific DEA methodology terms, linked with corresponding application areas as well as with some

geographical references. Also, the ego-network of some sample keywords is shown, and some research gaps in DEA are identified.

**Keywords:** DEA; keywords; co-occurrence network; complex network analysis; emerging topics; research gaps

## 1. INTRODUCTION

Data envelopment analysis (DEA) is a well-known methodology for the efficiency assessment of organizational units (OUs) (Cooper et al. 2004; Cooper et al. 2006; Zhu 2002). It is a data-driven non-parametric approach, which only uses the data on the inputs consumed and the outputs produced by each OU. DEA is a very active research field with hundreds of papers published on the topic each year, some of them more theory/methodology-oriented and others more applications-oriented. Providing insight into the complex web of relationships between the concepts, models, variants and extensions, applications, etc. that constantly appear in DEA seems a worthy endeavour, one that researchers and practitioners should welcome.

Several surveys of the DEA literature have been carried out over the years (e.g. Emrouznejad et al. 2008; Liu et al. 2013b). The most recent one (Emrouznejad and Yang 2018) reports the 50 most used keywords by number of DEA-related articles as well as the most popular keywords in the last two years included in their study (2015 and 2016).

There have been several papers dealing with the bibliometric analysis of the DEA literature using network analysis (NA) tools. Thus, Lampe and Hilgers (2015) carried out a co-citation analysis of the of DEA literature as well as of that of a parametric sister methodology (Stochastic Frontier Analysis, SFA). They made clear the need for data cleaning to clean out misspellings, match abbreviations, normalize the letters (capital/small letters), etc. The aim was to identify clusters of highly connected papers presumably dealing with similar topics. Thus, they found clusters associated with different economic sectors (e.g. electricity generation plants, telecommunications, airports, agriculture, energy efficiency, environmental performance, supplier selection, forestry, fishery, urban transit systems, etc.) as well as methodological clusters (e.g. network DEA, returns to scale, imprecise data, fuzzy set theory, ranking of units, super-efficiency, etc.). They also provided a citation analysis to identify seminal and must-read papers and the evolution of citation patterns over time.

Liu et al. (2016a) built a DEA citation network and identified four main research fronts in DEA, namely “bootstrapping and two-stage analysis”, “undesirable factors”, “cross-efficiency and ranking”, and “network DEA, dynamic DEA and SBM”. These research fronts were found using the Girvan and Newman (2002) clustering algorithm which removes arcs in decreasing order of their edge-betweenness and retains the clusters’ structure that maximizes the corresponding modularity function. Also, some key citation paths are identified and show the historical development of the field through key paper milestones. This is done using Main Path Analysis.

Liu et al. (2013a) also used Main Path Analysis to obtain local and global main paths in DEA. Not surprisingly, all the main paths start with the seminal work of Charnes et al. (1978). Also, by exploring secondary paths, different branches of DEA literature can be detected. In particular, Liu et al. (2013a) identified five active DEA subareas, which they labelled “two-stage contextual factor evaluation framework”, “extending models”, “handling special data”, “examining the internal structure” and “measuring environmental performance”. In this citation analysis study, the Web of Science (WOS) database was used as the data source. The papers were identified using certain query terms, such as ‘DEA’, ‘data envelopment analysis’, ‘Malmquist index’, ‘constant returns to scale’, ‘variable returns to scale’, ‘non-parametric efficiency’, ‘Farrell efficiency’, etc. These terms were searched for in the title, abstract and author keywords. The existence of many variations on these terms was noted (e.g. Malmquist indices’, ‘Malmquist indexes’, ‘data envelopment model’, ‘data envelope analysis’, ‘data enveloping analysis’, etc.) as well as papers that contain the acronym DEA but not with the meaning of Data Envelopment Analysis.

Ho and Liu (2013) applied NA to a DEA co-authors network, identifying author’s location, seniority, h-index, research discipline, etc. and using regression analysis to test these factors in explaining their brokerage capabilities in diffusing knowledge across regions and continents. More recently, Yu et al. (2017) have carried out a clustering of the DEA literature using a similarity coefficient based on co-citation and co-word analysis, identifying from the title and abstract of the papers that belong to each cluster the main research topics addressed in those papers.

From the above comments, it can be seen that most NA bibliometric applications to DEA are based on co-citation networks (in which the nodes are the DEA papers) or co-authors network (in which the nodes are the DEA authors). In some of those

approaches from the clustering of the papers different subareas within the field are identified and from the metadata (title, abstract and keywords) of the corresponding papers certain labels describing those clusters of papers are obtained. However, the relationships between the papers keywords are not captured or analysed. Those relationships contain valuable information about the knowledge structure of the field, i.e. its relevant concepts and their connections.

In this paper a different approach is used. The rationale is that the pattern of connections between co-keywords reflect their respective roles and importance and that this can be captured by using a NA approach. This methodology can provide insight about the whole network (e.g. identifying its community structure or detecting its small world character, assortativity, cliquishness, etc) and about the different nodes (e.g. their strength, centrality, neighbourhood). Moreover, taking into account the year of publication of the different papers the evolution of the relationships between the concepts can be modelled and hence the dynamics of the DEA knowledge structure and research priorities. This type of analysis has not been done before in the DEA field. Therefore, the idea is to build a network of keyword co-occurrences within a broad sample of DEA papers. The nodes in this network are the keywords that appeared in those papers, keywords that need to be pre-processed in order to remove all misspellings, variations, abbreviations, upper/lower case letters, etc. issues). The network is undirected and weighted. The weight of each link is the number of papers in which the two keywords appear together. The larger this weight the stronger the association between those terms. Using NA tools to study the DEA keywords co-occurrence network (looking, for example, for clusters of terms or studying the evolution of the network with time) is the purpose of this paper.

The structure of the paper is the following. Section 2 reviews the literature on keyword NA. In Sections 3 and 4 the methodology used and the results obtained are presented, respectively. Finally, Section 5 summarizes and concludes.

## **2. LITERATURE ON KEYWORD NETWORKS ANALYSIS**

Although networks, and graphs in general, have always been a useful modelling resource in Operations Research and Management Science, it has not been until the last decades of the last century that the systematic study of the structure and properties of the complex networks that can be found in the real world has attracted attention. This

has given rise to a new field in its own that can be labelled complex networks analysis, social networks analysis or, simply, networks analysis (NA).

The realization that real-world networks did not fit the characteristics of random graphs but showed certain interesting features, such as clustering (i.e. the neighbours of a node are often connected among them), small world (i.e. average path lengths increase at a much lower rate than the number of nodes), scale-free (i.e. the degree of distribution follows highly skewed power-law distributions) or assortative mixing (i.e. nodes with certain attributes are often connected with nodes that have similar attributes), has greatly enhanced our understanding of those networks. Topics such as detection of communities, vulnerability of networks to random failures and intentional attacks, cascading failures, diffusion and contagion processes on networks, etc. have received much research attention and produced significant developments. The interested reader is referred to Newman (2003) for an introduction to the theory and to da Fontoura Costa et al. (2011) for an overview of applications. The latter span all sectors: social networks, world wide web, transportation networks of all kinds (road networks, air transportation, railways, maritime transportation, etc.), financial networks, supply networks, trade networks, power grids, collaboration networks, tourism destination networks, etc.

Among the many applications of NA, it has been used for bibliometric analysis. Usually those bibliometric networks consist of nodes, which can be journals, researchers, keywords or publications, and links, which represent the relationships among them. Complex Network Analysis, CNA, allows performing a bibliometric research based on scientific collaboration using co-authorship relations, citations using citation relations or co-cited relations, keyword co-occurrence relations, among others (Van Eck and Waltman 2014). In each case, the corresponding network represents the set of papers under study using a different approach. Co-citation analysis, for example, studies the importance of each paper for the scientific community. Co-authorship analysis shows the researchers who have collaborated (i.e. co-authored) papers on certain topics. *Co-word analysis* studies the relationships among the used keywords to reveal the structure and development of each particular application or methodology.

The words needed to create a co-word analysis can be collected from the titles of the papers, as in Milojević et al. (2011) or from the abstracts and author keywords as in Zhao et al. (2018) and Choi et al. (2011). These different approaches allow us to identify subareas in each field and study their characteristics and trends (Lee et al.

2018), portray the global research profile (Liu et al. 2016b; Xu et al. 2016), find hot topics (Zhao et al. 2018), disruptive trends (Dotsika and Watkins 2017), look for cooperative relationships and interpret collaboration patterns among authors (Newman 2001a; Newman 2001b), etc.

Different ways to carry out bibliometric analysis have been proposed. One of those methods represents the relatedness of the keywords and employs distance-based visualization techniques, such as multidimensional scaling (MDS), to describe the knowledge structure of the field (e.g. Cho 2014; Yan et al. 2015). Other methods use clustering algorithms. For instance, in the case of co-word networks, Milojević et al. (2011) use hierarchical clustering based on the Jaccard coefficient; Yan et al. (2015) use hierarchical clustering using Ward's methods; Delecroix and Epstein (2004) use an Ascendant Hierarchical Clustering algorithm based on the strength of association with cluster keywords; and Williams et al. (2016) define clusters with the Louvain community detection method (Blondel et al. 2008); de la Hoz-Correa et al. (2018) and Godwin (2016) use Clauset et al.'s (2004) community detection algorithm and, in the case of citation networks, Lee et al. (2018) cluster the relationships among the themes using edge-betweenness (Girvan and Newman 2002).

Another important method for bibliometric analysis uses the metrics, layouts and visualization capabilities of NA. NA metrics allow us to categorize bibliometric networks. Regarding topology, Gan and Wang (2015) and Choi et al. (2011) observe that keyword co-occurrence networks follow a power-law distribution. Thus, the networks that fit the preferential attachment assumption are scale-free, i.e. new nodes tend to link preferentially with highly connected nodes. In other words, highly frequent keywords are more likely to be used together with newly introduced keywords. Delecroix and Epstein (2004) use the average strength of association among keywords with a metric called the Association Index as a basic metric in their study. These authors stress the importance of standardizing and eliminating spelling differences and variations of the same terms.

Wood and Khan (2015) and Khan and Wood (2015) use different centrality measures, such as degree centrality, betweenness centrality and eigenvector centrality in co-word networks, in papers dealing with international trade negotiations and information technology management domains, respectively. Lee et al. (2018) use the keywords with high centrality to show the overall structure of human space exploration.

Liu et al. (2016b) work with co-citations networks to study research trends, using the pathfinder method in co-word networks to analyse hot topics in the magnetic nanoparticles field; Zhao et al. (2018) rank the keywords based on several node metrics over co-word networks of three different disciplines, and establish keyword frequency as an effective method to identify hot themes. Additionally, a positive correlation between frequency and centrality measures based on degree and betweenness is also observed in Choi et al. (2011). Apart from that, Dotsika and Watkins (2017), using the methodology of Liang and Chen (2011), claim that the disruptive trends belong to the group of keywords with low degree centrality and high closeness. More recently, Huang et al. (2019) have carried out a co-word analysis of the keywords of papers concerning industrial symbiosis, detecting the small-world character of the corresponding network, carrying out a node centrality analysis and studying its mesoscopic structure.

From the literature review it can be concluded that: 1) bibliometric analysis is a useful methodology to extract knowledge of the developments and relationships among the research topics in different disciplines with the advantage of being based on objective bibliometric data, 2) NA is one of the tools that have been successfully applied to bibliometric analysis, 3) although citation and co-citation analyses have been applied to DEA, a similar keyword co-occurrence analysis has not been carried out in DEA, 4) although co-keywords analysis generally involve papers published in a multiple year period, the temporal evolution and the dynamics of this type of network has not been studied, and 5) the conceptual neighbourhoods (ego-networks in the NA terminology) induced by the keywords network have not been analysed. The first two observations led us to choose NA as research methodology while the last three motivated the study carried out and guided its development.

In summary, efficiency assessment is a dynamic research field in which the corpus of concepts, models and techniques keeps growing every year and which therefore can benefit greatly from mapping all that knowledge through a study of the keywords that describe the research done in this field. Although previous efforts to study the structure of this discipline have focused on co-citation and co-authorship relationships, we believe that additional insight can be gained interpreting the co-occurrence of keywords as a signal of their being related/complementary and representing and analysing those relationships within the complex networks paradigm.

### 3. METHODS

Although DEA is field with a long history, dating back to 1978, in this paper we are interested in the most recent developments of this dynamic research field. Specifically, ten years was considered a reasonable time span to focus our attention on. In order to build the DEA keywords co-occurrence network, all the keywords in all DEA papers published in the period 2008-2017 were collected. This required identifying first the corresponding papers. For that purpose, Elsevier's Scopus database was chosen because of its completeness (with over 71 million records), ease of use, and the availability of Scopus Search API, which facilitated the automated extraction of the requested data. All the papers published in any journal covered by Scopus in the period 2008 to 2017, containing in the title, abstract or keywords the terms "DEA" or "Data Envelopment Analysis", were searched and their bibliometric data (title, journal, year, digital object identifier and list of keywords) extracted. A total of 5,290 papers were identified, with an increasing number of appearances along the period (from 298 papers in 2008 to 789 in 2017). Since in some papers "DEA" had a different meaning (e.g. Dielectric Elastomer Actuator), a manual exploration of all those records led to deleting 78 of them as not dealing with efficiency-related topics, thus keeping 5,212 papers for our analysis, containing a total of 9,329 different keywords.

A quick analysis of the raw keywords showed some of them to contain misspellings (e.g. "Analytic hierarchy processs"), American vs. UK spelling (e.g. "Demutualization" and "Demutualisation"), singular and plural variants (e.g. "Abatement cost" and "Abatement costs"), hyphens (e.g. "Casemix" and "Case-mix"), synonyms (e.g. "Air cargo" and "Air freight" or "Delphi", "Delphi method" and "Delphi technique"), acronym variants (e.g. "Analytic hierarchy process (AHP)", "Analytic hierarchy process", "Analytical hierarchical process (AHP)", "Analytical hierarchical process", "Analytical hierarchy process (AHP)" and "Analytical hierarchy processing (AHP)") and other characteristics that made it necessary to clean up and standardize the raw list of keywords. This required a manual, lengthy and careful examination of the records, amending those issues, renaming some keywords, grouping together similar ones (e.g. "Air pollution", "Air pollutants", "Air emissions" and "Air pollution abatement") and expanding some acronyms whose meaning might be clear from the context of the paper in which the keyword appeared but not when the keyword is separated from it (e.g. the keyword "APS" was substituted by "Agrarian productive



strategies (APS)”, the keyword “AID” was substituted by “Automatic interaction detector analysis (AID)” and the keyword “EKC” was substituted by “Environmental Kuznets curve (EKC)”). This often required accessing the individual papers in which these keywords appeared and looking for their meaning and context.

There were some generic keywords that did not provide useful information (e.g. “Analysis”, “Analyzing”, “Analytical methods” or “Analytical techniques”) and were thus removed. Note in this regard that the generic keywords “DEA” (and its equivalent forms, “Data envelopment analysis”, “Data Envelopment Analysis”, “DEA-based method”, “DEA-like model”, etc.) and “Efficiency” (and its equivalent forms, “Efficiency analysis”, “Efficiency assessment”, “Efficiency appraisal”, “Efficiency estimation”, “Efficiency evaluation”, “Efficiency calculation”, “Efficiency index”, “Efficiency measure”, “Efficiency measurement”, etc.) were removed because those two trivial nodes would be connected to almost all other nodes in the network without providing any interesting information.

Some compound keywords were split in two (e.g. “AHP/TCH Method” was split into “AHP” and “Tchebycheff method”, the keyword “CO2 emissions uncertainty” was split into “Carbon dioxide emissions” and “Uncertain data” and the keyword “Energy conservation and emission reduction” was split into “Energy conservation” and “Emission reduction”). Moreover, many specific, low frequency keywords were split linking them to a larger category. Thus, for example, “ATM” was split into “Banks” and “Automated teller machine (ATM)”, “Baseball” was split into “Sports” and “Baseball”, “average length of stay” was split into “Healthcare sector” and “Average length of stay”, “Black start” was split into “Power grid” and “Black start”, “Cap and trade” was split into “Emissions trading” and “Cap and trade”, etc. This was done to prevent the information provided by low frequency nodes (which correspond to highly specific keywords) from being completely lost if they were filtered out. Thus, in NA, nodes with degree one are called pendants or dangling nodes and are generally considered to be of lesser importance and sometimes filtered out. In that case, this mechanism of linking a specific keyword with a larger category keyword (which has a higher frequency) allows the retention of part of the information of the keyword filtered out.

The way the geographical-related keywords were treated must also be mentioned. First of all, it must be mentioned that those keywords were retained because in addition to the relationships between the different efficiency analysis concepts and

techniques, we are interested in investigating the geographical distribution/pattern of the corresponding applications of these techniques. For example, since there is a large community of DEA researchers affiliated to Chinese institutions, it would not be surprising that many DEA applications refer to companies, sectors or administrative regions in that country, something which is certain when the corresponding geographic reference is explicitly included as a keyword. We believe that this type of keyword information should not be lost and that it can help to analyse the field from this specific angle and see if there more DEA applications of a certain type (education, transport, health care, environmental efficiency, agriculture, etc) in certain countries or world regions than in others. A two-level (country/region) was considered so that those keywords that represented countries and had low frequency could increase their connectivity through their corresponding region keyword. Thus, for example, the keyword “Egypt”, which appeared three times, was split into two, “Egypt” and “Africa”. Thus, if a paper presented an application involving Egypt then automatically all the other keywords in that paper would also be connected with the keyword Africa. The same was done, for example, in the case of the keyword “Spain”, which was split into “Spain” and “Europe”. This two-level criterion was not applied in the case of the keywords “USA” and “China”, which were not split to refer their corresponding regions. Keywords representing names of cities or provinces were substituted by their country keywords (e.g. the keyword “Illinois” was substituted by “USA”, the keyword “Andalusia” was substituted by “Spain” and the keyword “Beijing” was substituted by “China”). Moreover, some keywords had an implicit geographical information which was made explicit. Thus, for example, the keyword “Spanish football clubs” was split into “Football clubs” and “Spain” and the keyword “Brazilian electricity power” was split into “Electric power industry” and “Brazil”. Two, more complex examples are: the keyword “CNX Energy Index of India” was split into three “Stock exchange index”, “Energy sector” and “India”, and the keyword “Changjiang MSA” was split into “Maritime transport”, “Maritime safety administration (MSA)” and “China”.

All this information is provided to show and justify the level of intervention, implied by the cleaning and standardizing of the keywords, carried out. Although this was done with sense and judgement by the authors, a certain level of subjectivity and discretion was inevitably introduced in the analysis during this process. In the end, the number of keywords were reduced from the original 9,329 keywords to “just” 5,979,

leading to 38,519 pairs of keywords co-occurrences (i.e., edges in the network, representing pairs of keywords appearing in the same paper). There is a giant component that contains 5,899 nodes and 38,446 links.

The NA software used includes the R packages *igraph* (Csardi and Nepusz 2006) and *poweRlaw* (Gillespie 2015) for the calculations of the characterization metrics and *Gephi* for the visualization of the network. Ten networks have been created to evaluate the evolution of the DEA keyword co-occurrence network over the period 2008-2017, each network covering the keyword co-occurrences from 2008 until the corresponding year.

#### 4. RESULTS AND DISCUSSION

Table 1 shows some general information about the 10 networks being analysed. As it can be observed, the density of the networks decreases over time as the number of new keywords added outnumbers the growth of the links between the existing and the new ones. This is also responsible for the decrease in the average clustering coefficient. However, the average degree and average link weight increase due to the existence of some highly connected nodes (keywords appearing in many papers), although most of the nodes have a low degree. The calculation of the average path length and diameter is based on the effective length (i.e. inverse of the co-occurrence frequency). While the diameter decreases steadily (because of a reduction of the link's effective length) the average path length fluctuates around 6.50. The disassortativity of the giant component shows a slight but steady increase over the years, which indicates that low connected nodes tend to connect to highly connected ones.

Table 1. Some characterization measures of the (2008-20xx) networks

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
# Nodes	641	1,259	1,762	2,302	2,917	3,501	4,025	4,611	5,249	5,979
# Links	2,107	4,797	7,584	10,743	14,694	18,768	22,743	27,045	32,287	38,519
Density (%)	1.03	0.61	0.49	0.41	0.35	0.31	0.28	0.25	0.23	0.22
Aver. degree	6.57	7.62	8.61	9.33	10.07	10.72	11.30	11.73	12.30	12.88
Aver. link weight	3.58	4.23	4.89	5.38	5.93	6.43	6.89	7.26	7.71	8.19
Aver. clustering coeff.*	0.31	0.19	0.15	0.13	0.13	0.12	0.11	0.11	0.10	0.10
Average path length*	7.17	8.00	6.25	6.00	6.50	6.22	5.98	6.33	6.33	6.66
Diameter*	3.57	3.53	3.35	3.28	3.24	3.19	3.17	3.14	3.11	3.09
Assortativity*	-0.077	-0.078	-0.086	-0.094	-0.095	-0.101	-0.103	-0.114	-0.122	-0.129

\* Giant component

The cumulative distribution of the degree (Figure 1) follows a power law below a threshold around a degree of 100 and a log-normal distribution above it. In any case, the distribution is highly skewed, indicating that most of the nodes have a small degree but a few nodes have a rather large degree. Figure 1 also shows the cumulative distribution of the strength and of the link weights (i.e. the co-occurrence frequencies).

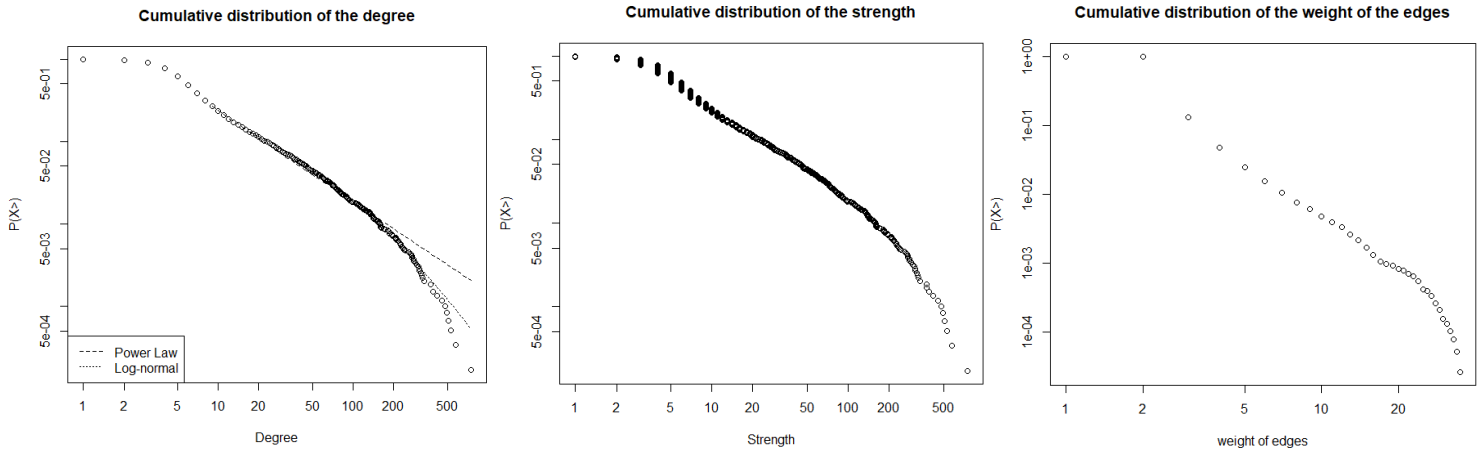


Figure 1. Cumulative distribution of the degree in the whole network (2008-2017)

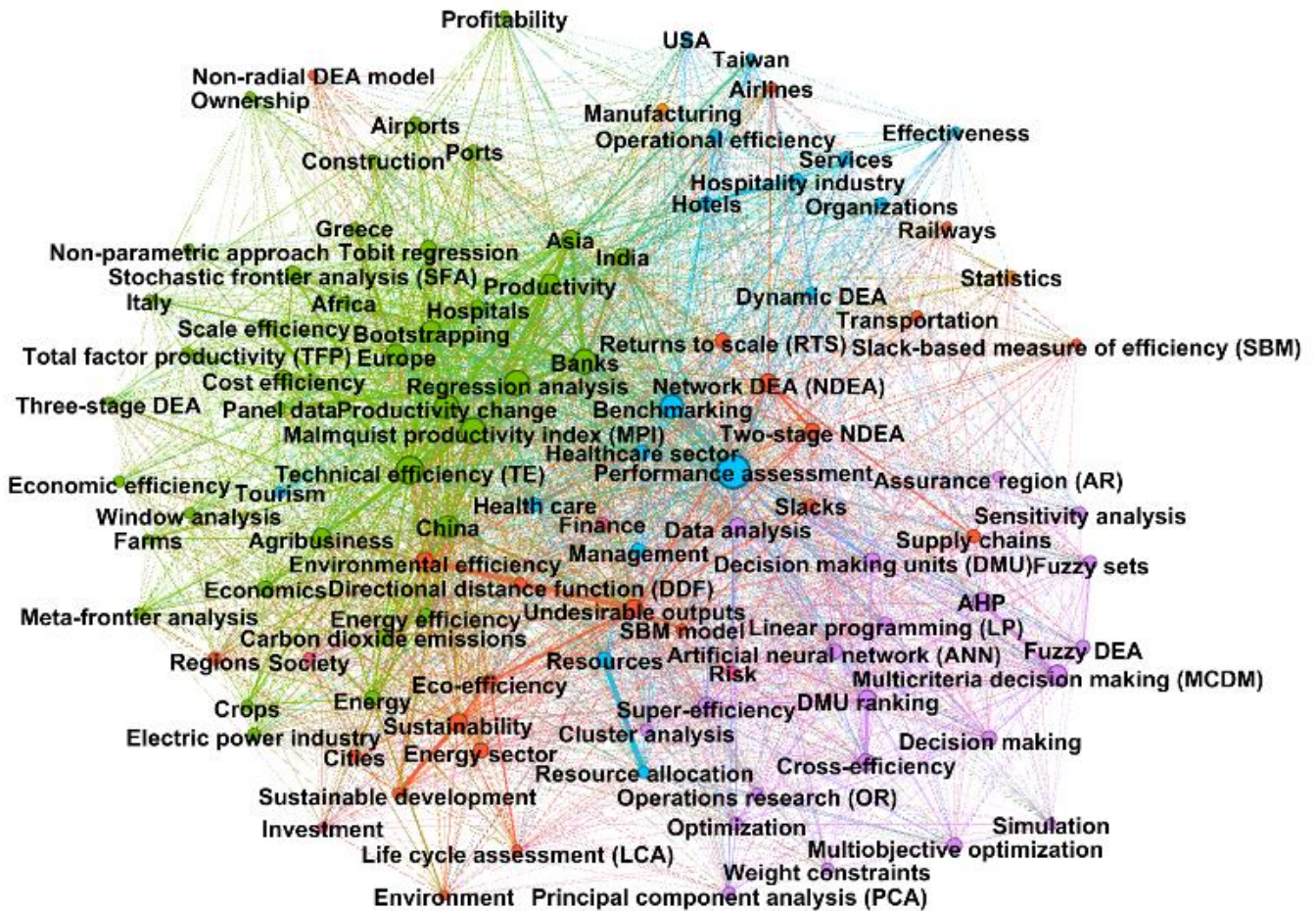


Figure 2. Largest k-core (k=30) of the 2008-2017 network

Figure 2 shows the largest k-core (corresponding to  $k=30$ ) of the 2008-2017 network (i.e. all keywords shown are connected to at least 30 of the other keywords shown). Note the existence of a strong connection between certain keywords, such as, for example, Environmental efficiency, Energy efficiency, Carbon dioxide emissions, undesirable outputs, Directional distance function. Several large communities formed by densely connected keywords loosely connected with other communities have been identified using the algorithm proposed in Clauset et al. (2004) and have been assigned different colours in the figure. Figure 3 shows the largest k-core within each of the five largest communities identified (sized 1,392; 1,341; 990; 879 and 376 nodes) which together represent 80% of the nodes. Each of these communities (from top to bottom and left to right) represents different groups of keywords:

- Operations Research keywords, including, for example, Simulation, Optimization, Multiobjective optimization, Metaheuristics, Artificial neural networks, Decision support systems (DSS), Goal programming (GP), Multicriteria decision making (MCDM), AHP, TOPSIS, Principal components analysis (PCA), Discriminant analysis, fuzzy DEA, etc.
- Second-stage and Three-stage regression, Productivity change, Malmquist and Malmquist-Luenberger productivity indexes, Cost and revenue efficiency, Meta-frontier analysis, Stochastic Frontier Analysis (SFA) and some application areas (Energy production, Greenhouse gases emissions, Ports, Airports, Agriculture, Telecommunications, Banks). This community contains also many geographical keywords (e.g. China, India, Japan, Iran, Turkey, European Union, Africa, MENA countries, ASEAN countries).
- Directional distance function (DDF), Undesirable outputs, Environmental efficiency, Slacks-based measures, Network DEA (NDEA), Game theory, Returns to scale (RTS), Negative data, Eco-efficiency and Sustainability, Supply chains, Electricity distribution, Food industry, Railways, etc. This community also contains some geographical keywords (e.g. Brazil).
- Dynamic DEA, Context-Dependent DEA and basically the service sector (Health care, Education, Hospitality industry, Tourism, Water utilities). This community contains also some geographical keywords (e.g. USA and Taiwan).
- Information technologies (IT), Information systems (IS), Project management, Research & Development (R&D), Financial performance, Bankruptcy, Science

& Technology (S&T), Innovation, Patents, High-tech industries,  
Manufacturing, Oceania, Pacific region, etc.



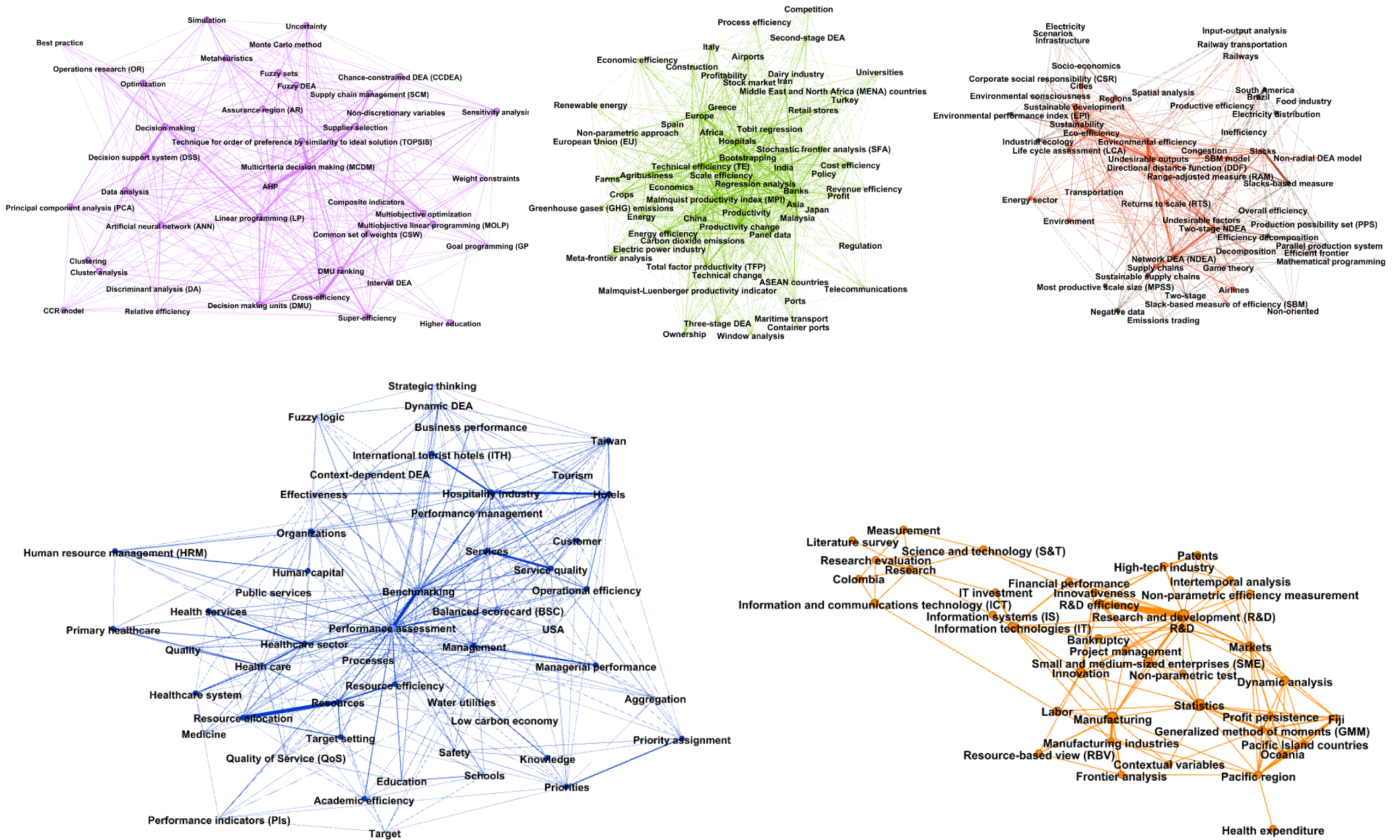


Figure 3. Largest k-core of the five largest communities of the giant component of 2008-2017 network

Although the subgraphs shown in Figure 3 are just a fraction (the most connected keywords) of the corresponding community, they are enough to show many interesting relationships that are intuitive to the experienced DEA researcher. For example, it is logical and reasonable the close relationships between concepts like Undesirable outputs, Environmental efficiency, Eco-efficiency, Sustainability, Environmental consciousness, Life Cycle Assessment (LCA), Industrial ecology, Environmental performance index (EPI), Environment, Transportation, Energy sector, etc. The link between some of these concepts with others such as Directional distance function (DDF), Slacks-based measure of efficiency (SBM), NDEA and two-stage NDEA are due to the large number of papers that use these DEA models in these areas. Another example are the strong links between China and keywords like Energy efficiency, Carbon dioxide emissions or Productivity change, topics which are a priority as regards that country.

The clusters of DEA concepts identified in this study differ from those of Yu et al. (2017). This is not surprising since these researchers use a different methodology: first clustering papers, not keywords, and then identifying the main DEA topics studied in the papers of each cluster. Yu et al. (2017) acknowledge that the clusters they found are mostly related to the applications of DEA in different sectors. Nevertheless, there are certain similarities between the communities shown above and the clusters in Yu et al. (2017). Thus, for example, the community we have labeled “Operations Research and Statistics keywords” is almost identical to their “Cluster#1: Decision making analysis and fuzzy DEA”. Their “Cluster #2: Energy and Environment” is contained within our second largest community and both include China and other geographical references. Actually, our second largest community also include their “Cluster #5: Agriculture and farm”, “Cluster #6: Banking” and “Cluster #7: Transportation”. Their “Cluster #4: Public service” is similar to our fourth largest community, which also includes part of their “Cluster #3: Business company”. However, in the seven clusters of Yu et al. (2017) there is nothing equivalent to our third community (which groups together methodology concepts like DDF, NDEA, SBM or undesirable outputs with the environmental efficiency and related application areas) or our fifth community (centred around IT/IS, R&D, Innovation, Patents, High-tech industries, among others).



Figure 4 shows the giant component of the 2008-2017 network after deleting all the links having a frequency smaller than nine (and the isolated nodes resulting after that deletion). Only 0.48% of the links and 2.44% of the nodes (i.e. 144 nodes) remain. This represents the strongest associations between keywords. Some of those keywords correspond to geographical terms, others to application sectors and others to methodological concepts.

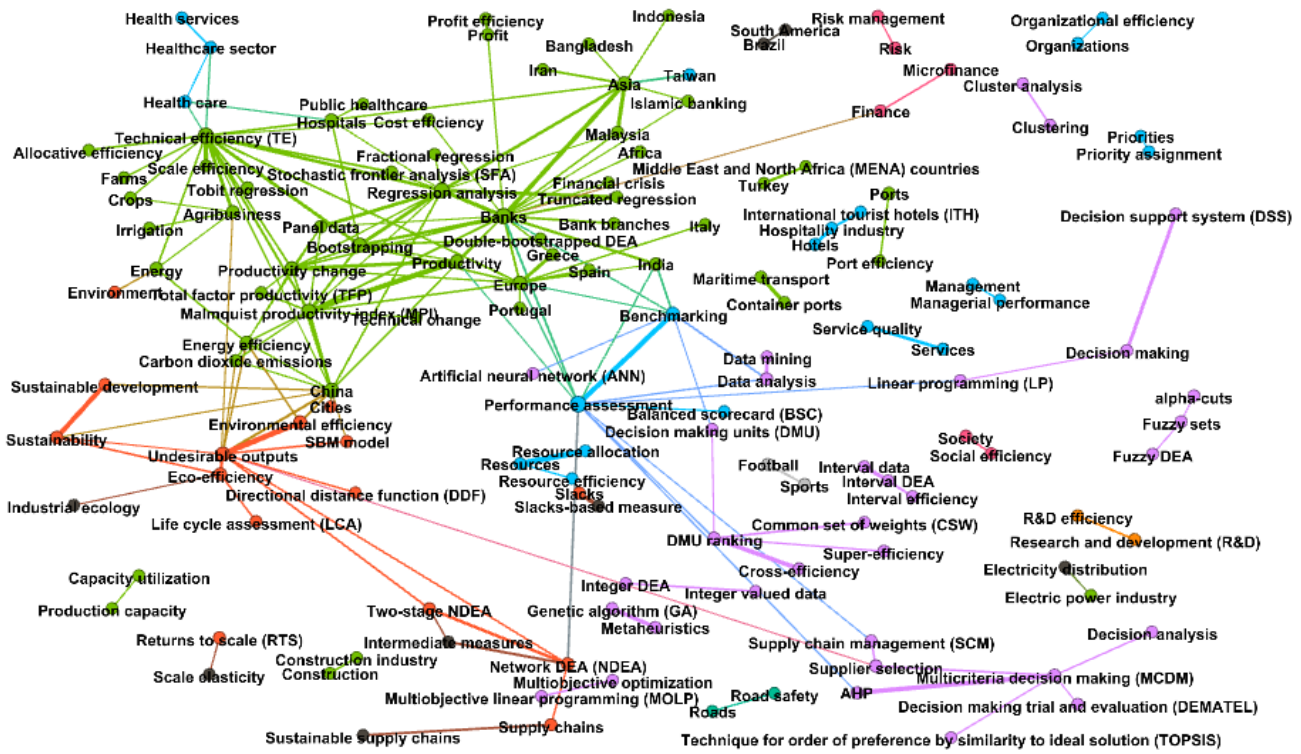


Figure 4. Giant component of 2008-2017 network (links with a weight lower than nine have been removed)

Many expected associations can be perceived in Figure 4. Thus, note the links between Technical efficiency and Scale efficiency and Allocative efficiency. Or the links between Profit and both Profit efficiency and Banks. Or between Scale elasticity and Returns to scale (RTS). Or between DMU ranking and Super-efficiency, Cross-efficiency and Common set of weights (CSW). Not unexpected either are the links between China and Sustainable development and between China and Environmental efficiency or between Supply chains and Network DEA (NDEA). The association between Banks and Asia and Banks and Europe is also interesting. But perhaps more important than the expected associations are those that are novel or unexpected. Thus,

for example, the high frequency of some keywords, such as Irrigation, Islamic banking or Balanced scorecard (BSC) is surprising. The abundance of Bootstrapping and Regression analysis keywords is also remarkable.

The interest of DEA researchers and the topics studied evolve over the years. New topics appear or gain interest and others lose interest. Table 2 shows the five keywords that appeared for the first time in a given year and that had the largest strength in the 2008-2017 network. These can be considered as the emerging topics in these years. The large presence of sustainability-related keywords in this table is remarkable (e.g. Green supply chain management (GSCM), Energy conservation, Environmental consciousness, Conservation, Sustainable suppliers, Sustainability efficiency, Sustainable tourism, and Sustainable operations). Uncertainty, Fuzzy DEA and Robust efficiency analysis are also significant. Other interesting emerging topics are Dual-role factors, Double-bootstrapped DEA, Dynamic analysis and Big data. Note also the presence of the SBM model as an emerging topic in year 2009, the same year that Tone and Tsutsui (2009) extended that model to NDEA.

Table 2. Keywords with the largest increment of strength from their first appearance

Year	Keywords
2008	Performance assessment (751); Technical efficiency (TE) (580); Banks (499); Regression analysis (512); Malmquist productivity index (MPI) (530)
2009	Agribusiness (310); Fuzzy DEA (209); Panel data (143); SBM model (141); Principal component analysis (PCA) (163)
2010	Uncertainty (124); Dual-role factors (62); Double-bootstrapped DEA (61); Academic efficiency (65); Green supply chain management (GSCM) (49)
2011	Learning (69); Energy conservation (68); Public healthcare (44); Technological gap ratio (TGR) (39); Input-output analysis (59)
2012	Islamic banking (42); Environmental consciousness (59); Dynamic analysis (54); Production function (40); Evolutionary algorithm (40)
2013	Priority assignment (43); Influencing factors (32); Mining (32); Water utilities (31); Private companies (25)
2014	Decision making trial and evaluation (DEMATEL) (36); Statistical methods (34); Development (40); Particle swarm optimization (PSO) (23); Conservation (27)
2015	Resilience engineering (40); Sustainable suppliers (23); Health sector reform (23); Sustainability efficiency (20); Regional industry (15)
2016	Big data (27); Sustainable tourism (20); County hospital efficiency (13); Urban development (21); Efficiency determinants (11)
2017	Robust efficiency analysis (20); Sustainable operations (16); Feasible generalised least squares (FGLS) (15); Regional efficiency (12); Meta-technology ratio (12)

Note: Degree in the 2008-2017 network shown within parentheses

Comparing the situation in 2008 versus 2017, Table 3 shows which the keywords with the largest strength were at the beginning and end of that period. Note that six of the keywords with largest strength in the 2008-2017 network were not present in the 2008 network. These are Sustainability, Undesirable outputs, Agribusiness, DMU ranking, NDEA and Super-efficiency. The last one is a remarkable feat for a rather specific DEA concept that seems to have endured after the limitations of some of its initial formulations were overcome. Out of the strongest keywords in the 2008 network, some of them, such as Performance assessment, Technical efficiency (TE), Malmquist productivity index (MPI) and Regression analysis, remained so in the 2008-2017 network. These keywords may thus be considered classical DEA keywords. Other keywords that ranked high in strength in 2008, such as Organizations, Stochastic frontier analysis (SFA), Resources, Economics or Tobit regression, have lost positions in this ranking. China has gained positions, while Europe has lost.

Table 3 Ranking of the 20 keywords with the largest strength in 2008 and 2017.

Rank in 2008		Rank in 2017
1	Data analysis	(21)
2	Malmquist productivity index (MPI)	3
3	Performance assessment	1
4	Technical efficiency (TE)	2
5	Banks	5
5	Europe	9
7	Regression analysis	4
8	Benchmarking	7
9	Productivity	14
10	AHP	12
11	Multicriteria decision making (MCDM)	10
12	Productivity change	6
14	Bootstrapping	10
14	Economics	(27)
14	Stochastic frontier analysis (SFA)	(35)
16	China	8
17	Asia	15
18	Tobit regression	(25)

19	Resources	(31)
19	Organizations	(83)
-	Sustainability	13
-	Undesirable outputs	16
-	Agribusiness	17
-	DMU ranking	18
-	Network DEA (NDEA)	19
-	Super-efficiency	20

A list of the 20 most important keywords according to different NA measures is shown in Table 4. The first four columns correspond to strength and centrality measures and the top ranked keywords are more or less the same. However, the keywords in the last column, which correspond to off-centre/marginal nodes, are completely different and seemingly less popular. Recall that degree centrality counts how many links a node has, without distinguishing that some links have more weight than others. Strength, in contrast, adds the weights of all the links connecting a node to its neighbours. In addition to taking into account the links weights, Eigenvector centrality considers the centrality of the neighbouring nodes: a node connected to central nodes is more “well-connected” (i.e. more central) than a node connected to “secondary” nodes. Closeness centrality averages the inverse of the effective geodesic distance from each node to every other node.

The effective distance between two connected nodes is the inverse of the weight of the link connecting them. Thus, if the weight of the link between two nodes is large then the effective distance is small, i.e. the nodes are intimately connected and are close. The geodesic distance between two nodes is the length of the shortest path between the nodes. The advantage of using the inverse geodesic distance is twofold: it works for nodes which are not connected (and whose geodesic distance is infinite) and it transforms distance into its opposite, closeness. The more central keywords, according to the measures considered, are basic DEA concepts such as Technical efficiency, Performance assessment, Benchmarking, Productivity change, Malmquist productivity index, Regression analysis, Bootstrapping, Undesirable outputs, Multicriteria decision making, Network DEA (NDEA). Banks, Mutual funds, Agribusiness, Hospitals and Energy/environmental efficiency are the most central applications areas, and, as regards geographical terms, Asia (especially China) and Europe stand out.

Table 4. Top keywords based on different NA measures

Rank	Strength	Degree centrality	Eigenvector centrality	Closeness	Eccentricity
1	(1320) Performance assessment	(751) Performance assessment	(82.51) Banks	(0.014) Performance assessment	(8) Marxian economics
2	(1109) Technical efficiency (TE)	(580) Technical efficiency (TE)	(79.75) Technical efficiency (TE)	(0.013) Banks	(8) Neoclassical economics
3	(1055) Banks	(530) Malmquist productivity index (MPI)	(78.23) Regression analysis	(0.013) Benchmarking	(8) Overinvestment
4	(981) Regression analysis	(512) Regression analysis	(72.74) Malmquist productivity index (MPI)	(0.013) Regression analysis	(8) Stagnation thesis
5	(927) Malmquist productivity index (MPI)	(499) Banks	(65.37) Bootstrapping	(0.013) China	(8) Surplus value
6	(798) Productivity change	(482) Productivity change	(65.07) Performance assessment	(0.013) Malmquist productivity index (MPI)	(8) Underconsumption
7	(778) China	(457) Benchmarking	(60.72) Productivity change	(0.013) Technical efficiency (TE)	(8) X-inefficiency
8	(767) Benchmarking	(418) China	(59.28) China	(0.013) Europe	(8) Semi-nonparametric regression
9	(726) Bootstrapping	(393) Europe	(53.15) Asia	(0.013) Network DEA (NDEA)	(7) Overproduction
10	(706) Europe	(376) Bootstrapping	(51.60) Europe	(0.013) Productivity change	(7) Probability-possibility
11	(706) Undesirable outputs	(376) Multicriteria decision making (MCDM)	(50.20) Productivity	(0.013) India	(7) Convex hull
12	(601) Asia	(338) AHP	(46.79) Benchmarking	(0.013) Bootstrapping	(7) Isoquant estimation
13	(596) Multicriteria decision making (MCDM)	(330) Sustainability	(44.66) Undesirable outputs	(0.012) Undesirable outputs	(7) Sum form
14	(520) Network DEA (NDEA)	(323) Productivity	(37.68) Environmental efficiency	(0.012) Supply chain management (SCM)	(7) Homothetic production function
15	(518) Agribusiness	(320) Asia	(35.91) Agribusiness	(0.012) Data analysis	(7) Homothetic hull
16	(513) Sustainability	(311) Undesirable outputs	(35.84) India	(0.012) Hospitals	(7) Enterprise architecture
17	(510) Productivity	(310) Agribusiness	(35.78) Tobit regression	(0.012) Mutual funds	(7) Fuzzy credibility constrained programming
18	(491) Environmental efficiency	(307) DMU ranking	(34.00) Hospitals	(0.012) Healthcare sector	(7) COBIT
19	(489) AHP	(298) Network DEA (NDEA)	(32.09) Panel data	(0.012) Asia	(7) Information technology management
20	(468) DMU ranking	(289) Super-efficiency	(31.88) Energy efficiency	(0.012) Bank branches	(7) p-Robustness

Note: The value of the corresponding NA measure is shown within parentheses





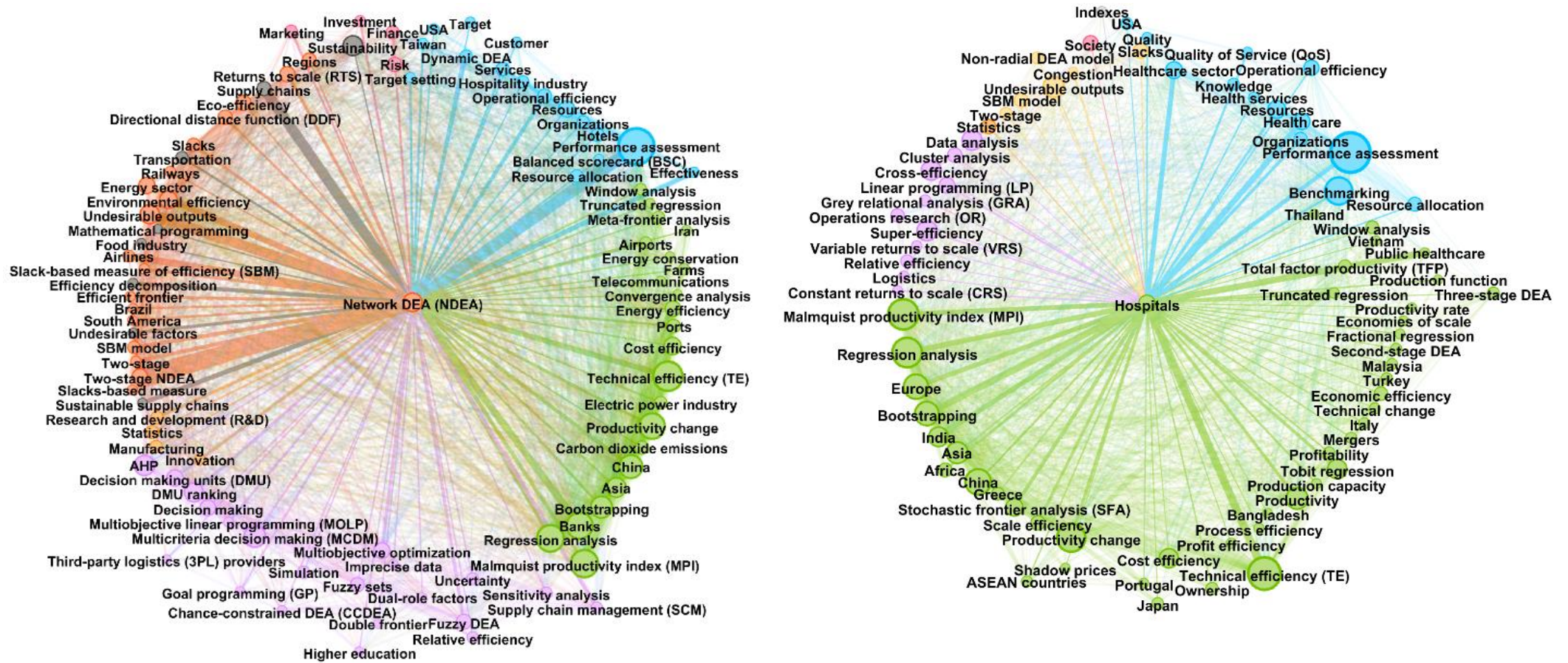


Figure 7. Ego-networks of “Network DEA (NDEA)” and “Hospitals”

As regards the ego-network of NDEA, that keyword has strong links with Supply chains, Two-stage NDEA, Undesirable outputs, Environmental efficiency, Dynamic DEA, SBM, DDF, RTS and Fuzzy DEA and with Airlines, Airports and Banks. Interestingly, it also has strong connections with Operational efficiency and Effectiveness.

Finally, the keyword Hospitals has strong links to Technical efficiency, Performance assessment, Malmquist productivity index, Regression analysis, Bootstrapping, Scale efficiency and SFA. As regards geographical associations, the strongest connections are with Europe and India. Surprisingly, it does not have strong links either with Cost efficiency or with Resource allocation. Connections with Centralised DEA, NDEA and Fuzzy DEA are also missing.

## **5. CONCLUSIONS**

This paper has carried out a NA of the co-occurrence network of the keywords appearing in the DEA papers published during the period 2008-2017. The relationships between those keywords reflect the knowledge structure of the field and its evolution along this period. Both the purpose of the analysis and the nature of the network used in this research differ from the co-citation and co-authorship analyses that have been applied before to study the DEA bibliography. Instead of considering a network of papers or a network of authors, a network of keywords is built from those of the corresponding papers. Thus, the co-word analysis carried out in this paper tries to address the fundamental questions of what the key concepts that make up the DEA field are, at both the theoretical and applications levels, and what their interrelationships are, i.e. how these concepts are logically connected and how they function together to add value to the research in the field. This type of analysis of the DEA bibliography has not been done before and provides new insight and a sort of conceptual map that can be useful for DEA practitioners and researchers alike.

Although the size of the keywords network steadily increases with time, the average path length does not, indicating that the DEA keywords co-occurrence network is a small world. It has also been observed that the degree, strength and weight distributions are highly skewed, with most nodes having low degrees and strength and most links having a low weight but with a few nodes with large degrees and strengths



and a few links with large weights. This indicates that researchers typically favour certain often-used, standard keywords that help to anchor the paper within a specific subarea but also use other less frequent terms to further delineate the specific theoretical or application context. Also, unlike what occurs in co-authorship networks, the DEA keywords co-occurrence network is disassortative, i.e. nodes with a low degree are generally connected with nodes with a higher degree and vice versa. This is consistent with the use, in a given paper, of both standard, high-frequency keywords and specific, low-frequency terms.

The node strength and other centrality indexes computed highlight the most relevant methodology keywords and application areas and their relationships. Although many of these relationships can be, often unconsciously, perceived by researchers in their study of the state of the art, applying NA helps make them explicit and visible, especially after the low frequency co-occurrences are filtered out, retaining the significant ones.

As regards the individual nodes, the more central DEA keywords are some common, generic and nondescript terms (such as Technical efficiency, Performance assessment, Benchmarking, Productivity change) as well as some frequently-used DEA concepts (Malmquist productivity index, Regression analysis, Bootstrapping, Undesirable outputs, MCDM and NDEA), certain applications areas (such as Banks, Mutual funds, Agribusiness, Hospitals and Energy/environmental efficiency) and some geographical references (such as China and Europe).

The NA methodology applied not only identifies the main relationships between pairs of keywords (microscopic view) but also reveals the existence of large communities of inter-linked keywords (macroscopic view). In particular, among the five largest community there is one mainly formed by Operations Research and decision-making keywords, another three that include both methodological terms and application areas (as well as geographical references) and another one centred around technology and innovation. These keyword communities reflect the specific knowledge structure of the DEA field and, although it can evolve with time, the keyword community structure is rather stable. Actually, the study of the network growth along the years shows the dynamics of this field, with the emergence of new research topics and other topics becoming mature or even declining in interest. This points to the existence of a keyword

life cycle pattern that would apply not only in this field but in all scientific areas. This short-term keyword life cycles are superimposed onto the corresponding long-term paradigm life cycle.

As an additional example of the possibilities of NA, the ego-networks corresponding to some of those keywords have also been shown. These ego-networks show in a visually appealing way the main associations of a given topic, thus revealing the main lines of research in relation to that topic as well as some missing links. The latter can be used to identify research gaps.

Therefore, as regards the contributions of this paper, on one hand, the NA carried out has provided a wealth of information, that cannot be found in the existing literature, on the key concepts involved in the DEA field, assessing their relevance, showing the differential evolution of their strength, identifying their closest-related concepts (i.e. their conceptual neighbours) and grouping them into meaningful communities of concepts. On the other hand, some of the findings of the current study (like small-world character, assortativity, clustering, power-law distribution, community structure, etc) have also been observed in other fields, and thus confirm the existence of general laws that govern the growth and diffusion of scientific knowledge.

Regarding the research gap addressed in this paper, the focus is on the network of relationships among the keywords in an important research scientific field such as DEA. It has been shown that the analysis of these relationships (instead of the most classical analysis of co-citation or co-authorship analysis) provide a novel perspective and a better understanding of the knowledge structure of the field and its dynamic character.

The main limitation of this study refers to the need for the raw keywords to be cleaned up and standardized, a process that improves the quality of the keywords but at the cost of partially modifying the original keywords and introducing subjectivity in the analysis. Another limitation is the limited time span covered and the need to keep the network updated with the new DEA papers that are published.

Regarding topics for further research, one of them would be developing and testing a keyword life cycle model which could be applied to the whole DEA keyword network or separately for each the communities identified. Also, link prediction

techniques can be used in this DEA keyword co-occurrence network to identify currently-unconnected research topics that can be reasonably coupled in the future. It would also be interesting to look for alternative ways of skeletonising the network and identifying its backbone. A core-periphery model can also be tested. Integrating the keyword co-occurrence network with existing text analysis techniques to automatically extract keywords from the title and abstract of the paper is also an interesting topic for further research. Finally, it is desirable to use this type of keyword co-occurrence network to aid researchers choose the most appropriate keywords for a specific study, thus helping standardize and keep track of the keywords used.

### **Acknowledgements**

This research was carried out with the financial support of the Spanish Ministry of Economy, Industry and Competitiveness, and the European Regional Development Fund (ERDF), grant DPI2017-85343-P. We are also grateful to two anonymous reviewers for their constructive comments and suggestions.

### **REFERENCES**

- Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10(P10008).
- Charnes, A., Cooper, W.W. & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6) 429–444.
- Cho, J. (2014). Intellectual structure of the institutional repository field: A co-word analysis. *Journal of Information Science*, 40(3), 386–397.
- Choi, J., Yi, S. & Lee, K.C. (2011). Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information and Management*, 48(8), 371–381.
- Clauset, A., Newman, M.E.J. & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70, 066111.
- Cooper, W.W., Seiford, L.M. & Zhu, J. (2004). *Handbook on Data Envelopment Analysis*. New York: Springer.

- Cooper, W.W., Seiford, L.M. and Tone, K. (2006). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd edition. New York: Springer.
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research, *InterJournal*. Volumen Complex Systems, 1695.
- da Fontoura Costa, L., Oliveira Jr, O.N., Travieso, G., Rodrigues, F.A., Ribeiro Villas Boas, P., et al. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60, 329–412.
- de la Hoz-Correa, A., Muñoz-Leiva, F. & Bakucz, M. (2018). Past themes and future trends in medical tourism research: A co-word analysis. *Tourism Management*, 65, 200–211.
- Delecroix, B. & Epstein, R. (2004). Co-word analysis for the non-scientific information example of Reuters Business Briefings. *Data Science Journal*, 3(15), 80–87.
- Dotsika, F. & Watkins, A. (2017). Identifying potentially disruptive trends by means of keyword network analysis. *Technological Forecasting and Social Change*, 119, 114–127.
- Emrouznejad, A., Parker, B.R. & Tavares, G. (2008). Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-economic Planning Sciences*, 42, 151–157.
- Emrouznejad, A. & Yang, G.L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, 61, 4–8.
- Gan, C. & Wang, W. (2015). Research characteristics and status on social media in China: A bibliometric and co-word analysis. *Scientometrics*, 105(2), 1167–1182.
- Gillespie, C.S. (2015). Fitting Heavy Tailed Distributions: The powerLaw Package. *Journal of Statistical Software*, 64(2), 1-16.
- Girvan, M. & Newman, M.E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99, 7821–7826.
- Godwin, A. (2016). Visualizing systematic literature reviews to identify new areas of research. *Proc. 2016 IEEE Frontiers in Education Conference*, 1–8, <https://doi.org/10.1109/FIE.2016.7757690>
- Ho, M.H-C. & Liu, J.S. (2013). The motivations for knowledge transfer across borders: the diffusion of data envelopment analysis (DEA) methodology. *Scientometrics*, 94 (1), 397–421.

- Huang, M., Wang, Z. and Chen, T. (2019). Analysis on the theory and practice of industrial symbiosis based on bibliometrics and social network analysis. *Journal of Cleaner Production*, 213, 956–967.
- Khan, G.F. & Wood, J. (2015). Information technology management domain: emerging themes and keyword analysis. *Scientometrics*, 105(2), 959–972.
- Lampe, H.W. & Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. *European Journal of Operational Research*, 240, 1–21.
- Lee, T.S., Lee, Y.S., Lee, J. & Chang, B.C. (2018). Analysis of the intellectual structure of human space exploration research using a bibliometric approach: Focus on human related factors. *Acta Astronautica*, 143, 169–182.
- Liang, Y. & Chen, J. (2011). Group Network Centrality Analysis of Blogs in Politics. *Communication in information science and management engineering*, 1(3), 32–35.
- Liu, J.S., Lu, L.Y.Y., Lu, W.M. & Lin, B.J.Y. (2013a). Data envelopment analysis 1978–2010: a citation-based literature survey. *Omega*, 41, 3–15.
- Liu, J.S., Lu, L.Y., Lu, W.M. & Lin, B.J. (2013b). A survey of DEA applications, *Omega*, 41, 893–902.
- Liu, J.S., Lu, L.Y.Y. & Lu, W.M. (2016a). Research fronts in data envelopment analysis. *Omega*, 58, 33–45.
- Liu, P., Chen, B.L., Liu, K. & Xie, H. (2016b). Magnetic nanoparticles research: a scientometric analysis of development trends and research fronts. *Scientometrics*, 108(3), 1591–1602.
- Milojević, S., Sugimoto, C.S., Yan, E. & Ding, Y. (2011). The Cognitive Structure of Library and Information Science: Analysis of Article Title Words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933–1953.
- Newman, M.E.J. (2001a). Scientific collaboration networks: I. Network construction and fundamental results, *Physical Review E*, 64, 016131.
- Newman, M.E.J. (2001b). Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality, *Physical Review E*, 64, 016132.
- Newman, M.E.J. (2003). The Structure and Function of Complex Networks, *SIAM Review*, 45(2), 167–256.
- Tone, K. & Tsutsui, M. (2009). Network DEA: A slacks-based measure approach, *European Journal of Operational Research*, 197, 243–252.

- Van Eck, N.J. & Waltman, L. (2014). Visualizing Bibliometric Networks, In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer
- Williams, R., Runco, M.A. & Berlow, E. (2016). Mapping the Themes, Impact, and Cohesion of Creativity Research over the Last 25 Years. *Creativity Research Journal*, 28(4), 385–394.
- Wood, J. & Khan, G.F. (2015). International Trade Negotiation Analysis: Network and Semantic Knowledge Infrastructure. *Scientometrics*, 105(1), 537–556.
- Xu, X., Wang, W., Liu, Y., Zhao, X., Xu, Z. & Zhou, H. (2016). A bibliographic analysis and collaboration patterns of IEEE transactions on intelligent transportation systems between 2000 and 2015. *IEEE Transactions on Intelligent Transportation Systems*, 17(8), 2238–2247.
- Yan, B.N., Lee, T.S. & Lee, T.P. (2015). Mapping the intellectual structure of the Internet of Things (IoT) field (2000–2014): a co-word analysis. *Scientometrics*, 105(2), 1285–1300.
- Yu, D., Wang, W., Zhang, S., Zhang, W. & Liu, R. (2017). Hybrid self-optimized clustering model based on citation links and textual features to detect research topics. *Plos One*, 12(10), e0187164.
- Zhao, W., Mao, J. & Lu, K. (2018). Ranking themes on co-word networks: Exploring the relationships among different metrics. *Information Processing and Management*, 54(2), 203–218.
- Zhu, J. (2002). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Boston, USA: Kluwer Academic Publishers