Co-Authorship of scientists in the energy field: an exploratory study of the ETDE World Energy Database (ETDEWEB) using Social Network Analysis*

Vasco Monteiro[1], Miguel Neto[1], Aurora Cardoso[2], Ana Maria R. Correia[3]

[1] Instituto Superior de Estatística e Gestão de Informação – Universidade de Lisboa, Lisboa, Portugal
[2] DEECA, INETI – Instituto Nacional de Engenharia, Tecnologia e Inovação, I. P., Lisboa, Portugal
[3] ISEGI – Universidade Nova de Lisboa, Lisboa, Portugal e IAPMEI – Instituto de Apoio às Pequenas e Médias Empresas e à Inovação, I. P.
m2007120@isegi.unl.pt
mneto@isegi.unl.pt
aurora.cardoso@ineti.pt
acorreia@isegi.unl.pt and ana.correia@iapmei.pt

ABSTRACT

This paper presents a preliminary exploratory study, using a social network analysis (SNA) approach to examine the structure of co-authorship collaboration within the research community in the energy field from 1995 to 2008. The domain of the study is Portuguese scientists, working either in Portugal or abroad; by foreign scientists working in Portugal or by scientists who have co-authored with either of these groups.

The study uses the most common measures of macro (whole network) and micro (actor-centered) structures of this collaboration.

The data used to design the social network was obtained from the *Energy Technology Data Exchange* (*ETDE)'s Energy Database*, which is the largest collection of energy research and technology literature in the world created under the umbrella of the *International Energy Agency/Organisation for Economic Co-operation and Development (IEA/OECD)*.

Keywords: Co-authorship, Collaboration, ETDEWEB, Social Network Analysis, SNA, Energy, Portugal, R&D

1. Introduction

In contemporary society, collaboration in research and scientific publication is very common in most areas of academic science. In fact, as Li-Chun et al. (2006) pointed out there is evidence that cooperation among researchers is increasing in all scientific fields.

As Hara et al. (2003) highlighted, in science and technology the effect of this cooperation is significant when addressing complex problems in the contemporary world of "rapidly changing technology, dynamic growth of knowledge and highly specialized areas of expertise".

Collaboration may be seen as a process where knowledge flows among scientists and through which individual scientists gain access to new "knowledge capital". When that collaboration widens, scientists may gain access to information both directly and indirectly; this may affect scientific productivity, both in quality and quantity (Li-Chun, Y et al., 2006).

---

* Monteiro, Vasco; Neto, Miguel; Cardoso, Aurora & Correia, Ana Maria (2009). Co-Authorship of scientists in the energy field: an exploratory study of the ETDE World Energy Database (ETDEWEB) using Social Network Analysis. Paper presented at *ECEMEI-5 - 5th European Conference Economics and Management of Energy in Industry* (14-17th April 2009, Vilamoura, Portugal)

During the first half of the twentieth century, scientific papers written by more than one author were relatively rare. However, as Acedo et al. (2006) note, this situation has changed significantly over the last decades.

This paper presents an exploratory study, using a social network analysis approach to examine the structure of co-authorship collaboration within the Portuguese research community, working in the energy field, from 1995 to 2008; it uses the most common measures of macro (whole network) and micro (actor centered) structures of this collaboration (Li-Chun, Y et al., 2006).

The data used to build the social network was obtained from the *Energy Technology Data Exchange (ETDE)'s Energy Database (*ETDEWEB*)*. By December 2008, there were over 8,000 records of publications involving scientists and research institutions in Portugal, covering the above-mentioned period. Based on the knowledge network constructed, we can analyse specific paths through which knowledge sharing occurred and by which knowledge capital was nurtured within the energy field in Portugal.

2. Co-Authorship and Social Network Analysis

The interest in the nature and scale of scientific collaboration, including co-authorship, is growing, especially in the ways that knowledge creation and sharing processes unfold (Barabási et al, 2002; Barnett et al., 1988; Katz and Martin, 1997; Laband and Tollison, 2000; Moody, 2004; Newman, 2001).

One of the possible approaches to assess scientific collaboration is based on the assumption that co-authorship creates a social network of researchers that develops over time (Hara et al., 2003). If one maps the network of co-authorship, using social network analysis (SNA), it is possible to infer the structure of the collaboration that is taking place between the network members.

This use of social network analysis as a methodology is finding increasing applications outside the social sciences and up to date it has been applied to areas as diverse as business organizations, electronic communication, health and psychology (Clark, 2006).

In networks of research communities, information and intellectual capital are among the most important resources; their flows and bottlenecks, within the communities, can be studied by social network analysis (Neto et al., 2008).

In theory, these studies should provide a basis for thinking about how a community is organized and what actions might be appropriate to create and develop an environment in which collaboration research is encouraged and ideas shared (Vidgen et al., 2007).

As Cronin (1996) points out, although the study of co-authorship relationships is only one of the possibilities for measuring formal and informal collaborations of scientists, it assumes particular relevance because it is fundamental in leveraging scientific activity; it can also be used as one of the tools to develop a reward system for academics.

This type of metric assumes an important contemporary role, because of the institutional pressure on researchers to increase their productivity by publishing scientific work and the fierce competition for the finite space available in scientific journals. "Publish or perish" dictates success or failure in the competition for funds and other resources (Piette, M. J. & Ross K. L., 1992).

3. The ETDEWEB database

The *ETDE – Energy Technology Data Exchange*, an international energy information exchange agreement, was formed in 1987 under the *International Energy Agency/Organisation for Economic*

Co-operation and Development (IEA/OECD); the Portuguese Government has designated the former *INETI (Instituto Nacional de Engenharia, Tecnologia e Inovação, I. P.)* as the Contracting Party for Portugal. The aim was to provide added visibility to the research work performed in Portugal by the research community within the scope of the *ETDE* ´s database. An Internet tool – the ETDEWEB (https://www.etde.org/etdeweb/) – was developed to disseminate the energy research and technology information, collected by 16 *ETDE* Member Countries and international partners, to all its other members; beginning in 2004, this free internet access has been extended to more than 50 developing countries worldwide.

ETDEWEB is the largest collection of energy research and technology literature in the world with a growing total of over 4 million abstracted and indexed records in the full collection. *ETDE* began the database in 1987 but historical energy-related information from the U.S. Department of Energy's Office of Scientific and Technical Information (OSTI) databases is also included, giving coverage back to 1974.

Input to the ETDEWEB is a shared endeavour; the entries are prepared by all 16 participating countries. Each member country is responsible for the input of the relevant publications produced within its frontiers. *ETDE* also has partnerships with other international organizations, which contribute to enrich the database content. In Portugal, the input is by *INETI*; the referees are *INETI´s* Researchers, and the database can be accessed by other Portuguese entities, on completion of a preliminary registration form (https://www.etde.org/etdeweb/register.jsp).

The ETDEWEB users include researchers, policymakers, academics, information specialists, and private citizens requiring answers to energy-related science and technology questions. They can identify latest developments, people and countries involved in a particular research area and energy-related environmental and climate change issues, including policy and economic factors, alternative and renewable energy sources and conservation.

The database contains bibliographic references and abstracts to journal articles, reports, conference papers, books, websites, and other miscellaneous document types (not commercially available). It also provides direct download access to 241,000 full text items via the *ETDE* operating site via the *OSTI* site, to 867,000 DOI links through which the item can be obtained from the publisher and to many more documents that are stored on organizations' websites around the world (as of March 2009).

The subject areas covered in the database are extensive; they are organized in nearly 50 subject categories. Some of the main areas include information on energy R&D; energy policy and planning; basic sciences (*e.g.*, physics, chemistry and biomedical) and materials research; the environmental impact of energy production and use, including climate change; energy conservation; nuclear (*e.g.*, reactors, isotopes, waste management); coal and fossil fuels; renewable energy technologies (*e.g.*, solar, wind, biomass, geothermal and hydro).

Subject categories are used by *ETDE* to classify records. Typically, these fall into four general types:

- those representing energy sources, *e.g.*, Coal, Lignite and Peat, Solar Energy, Wind Energy;
- those representing energy production, utilization, and management, *e.g.*, fossil-fueled power plants and energy conservation, consumption and utilization;
- those representing energy conversion and storage, *e.g.*, direct energy conversion and energy storage;
- those containing basic information developed in support of energy production, conversion, and utilization, *e.g.*, chemistry and physics.

Each record in the database has the following fields: title, creator/author, publication date, resource/document type, resource relation, size/format, subject, description/abstract, publisher, country of publication, language, source, availability, OSTI identifier, publication date and other identifiers. Records which describe documents that are in a native, non-English language will almost always include an English title and abstract to facilitate searching and help users to determine if translation is worthwhile.

4. Research Approach and the ETDEWEB data

For the present study, we used all registers from the ETDEWEB database involving Portuguese scientists or Portuguese research institutions. In this way one gets all the research that is performed in Portugal, either by Portuguese or foreign scientists. Based on the knowledge network constructed, one can analyse specific paths through which knowledge sharing occurred and by which knowledge capital was nurtured, within the energy field in Portugal.

Of over 8,000 existing records, we only considered those that were published between 1995 and 2008, where the subject field was completed, the resource type was Conference or Journal Article and where the creator/author was different from “None” or “Not Available”. With these constraints, 6,213 records were available for this study.

As Moody (2004) states, the generally lower rate of co-authorship in books may offset some of the errors introduced by their exclusion in this type of analysis. Consequently, books, technical reports and other miscellaneous items were not considered for the present study.

As a preliminary step in any SNA, a detailed data analysis of the ETDEWEB data was carried out. For instance, the field “creator/author” has both author’s name and affiliations; however authors’ names are shown in various ways, *e.g.*: surname, initial of the first name; initial of the first name and surname; name and surname, etc. Similarly affiliations can appear in full or in a variety of abbreviations; also an author can appear with one or two initials or indicate different affiliations in different publications. An actor/institution dictionary was created to deal with these variations and all records were updated accordingly. When it was not possible to allocate an institution to an author, the institution field was designated “Others”.

This option prevents the same author appearing on several nodes in the network, and introducing erroneous results.

5. Preliminary Results

After doing the split of the “creator/author” field, in order to divide the authors and the institutions, we had 73 new fields that needed to be processed. As a methodological approach, we looked first at the records that had the least number of completed fields.

With this approach, the preliminary results are based on 1,612 records of the ETDEWEB database,

Of these records more than 73% refer to journal articles, as indicated in Table I.

Table I Labour Type	#	%
Conference	424	26,3%
Journal Articles	1188	73,7%

As shown in Table II, the “Basic information developed in support of the energy field” represents almost 75% of the database; followed by 15% on research into “energy sources”; the rest concern research into “energy production, utilization and management” and finally research on “energy conversion and storage”, which only represents 1.2% of the total research in this sample.

Table II: Subject Comparison	#	%
Energy Sources	243	15,1%
Energy Production, Utilization, and Management	148	9,2%
Energy Conversion and Storage	20	1,2%
Basic information developed in support of energy	1201	74,5%

Table III illustrates the distribution of scientific literature production over the years. Years 2002-2007 have the highest scientific production, with more than 130 records per year. Year 2008 is the year with fewest records, in large measure this is because the data entry of 2008 journal articles or conference proceedings referring to this year will not be completed until 2009.

Table III: Scientific Production per year	#	%
1995	92	5,7%
1996	79	4,9%
1997	71	4,4%
1998	74	4,6%
1999	90	5,6%
2000	96	6,0%
2001	85	5,3%
2002	130	8,1%
2003	162	10,0%
2004	194	12,0%
2005	141	8,7%
2006	184	11,4%
2007	157	9,7%
2008	57	3,5%

Table IV highlights that 24,1% of the resources are single author publications and that publications by two authors are the most common type of co-authorship.

Table IV: Number of Authors per Resource	#	%
1	389	24,1%
2	397	24,6%
3	359	22,3%
4	198	12,3%
5	102	6,3%
6	59	3,7%
7	45	2,8%
8	23	1,4%
9	10	0,6%
>= 10	30	1,9%

Table V shows that the average number of authors per resource, since 1995 to 2008, is always higher than 2 authors per paper. The last 5 years, the average is higher than 3 authors, except in the year 2007.

Table V: Average Number of Authors per Resource (Per year)	
1995	2,54
1996	2,24
1997	3,23
1998	2,76
1999	2,69
2000	2,67
2001	3,09
2002	2,91
2003	2,98
2004	3,46
2005	3,48
2006	3,46
2007	2,82
2008	3,33

6. The Application of Social Network Analysis to the ETDEWEB

Based on the ETDEWEB data, processed as briefly described in section 4, a social network preliminary analysis of the co-authorship relationships, since 1995 to 2008, within the energy field, among scientists working in Portugal and Portuguese scientists having affiliations in foreign R&D institutions was carried out. The social network maps were built, using the *NetDraw* application (Borgatti, 2002), a free software tool for displaying social networks.

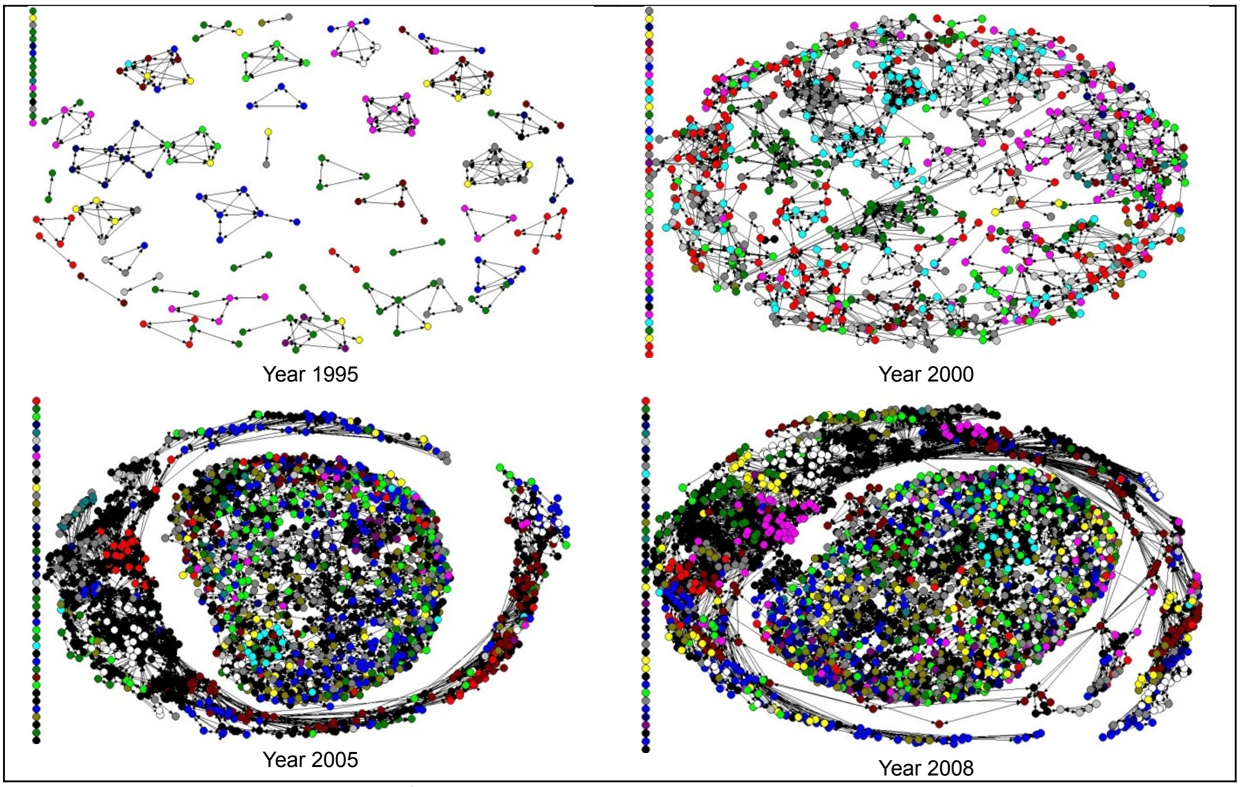

Figure 1: Energy field community co-authorship social network evolution

Figure 1 graphically illustrates the evolution of the social network in terms of actors (authors) and the relationships they have established. Each year, the social network map represents the cumulative co-authorships from 1995 up to the reference year.

The first conclusion that one can draw is the increase in co-authorship, which validates the trend mentioned by various authors (Acedo et al, 2006) and validates the previously noted tendency to cooperation when addressing complex problems in the contemporary world (Hara et al, 2003).

The nodes represent the social network actors, *i.e.* the scientists of the energy field working in Portugal and those whose affiliation is abroad. In this case, the nodes show the main institution attribute, based on the node colour and the lines represent the co-authorship relationships. Although the colour represents the institution, we can not make the comparison here between the years because of the large number of scientists (2440) and of institutions (166) which are represented.

One can further analyse some specific paths, through which collaboration occurred, using some of the centrality measures.

Centrality is a structural attribute of nodes in a network; using this measure, we can get some idea of the importance, influence and prominence of an actor in the network. According to Kilduff and Tsai, (2003) one can determine the relative importance of an actor by examining one of the following network characteristics:

- has many ties to other actors – degree centrality;
- is able to reach many actors – closeness centrality;
- connects to other actors who have no direct connections – betweenness centrality;
- is connected to many actors who are themselves connected to many other actors – eigenvector centrality.

Cumulative co-authorship relationships, from 2005 until 2008, were used in the centrality analysis.

Figure 2 illustrates the co-authorship social network, where the node size represents the degree of centrality and indicates the number of links incident upon a node. This indicator does not take into account the strength of direct links between the two actors; it only reveals the number of people with whom any one scientist has collaborated. This represents access to information and can be considered as a hub or a connector in this network. In Figure 2, we see that only a few scientists on the top left have this high degree of centrality.

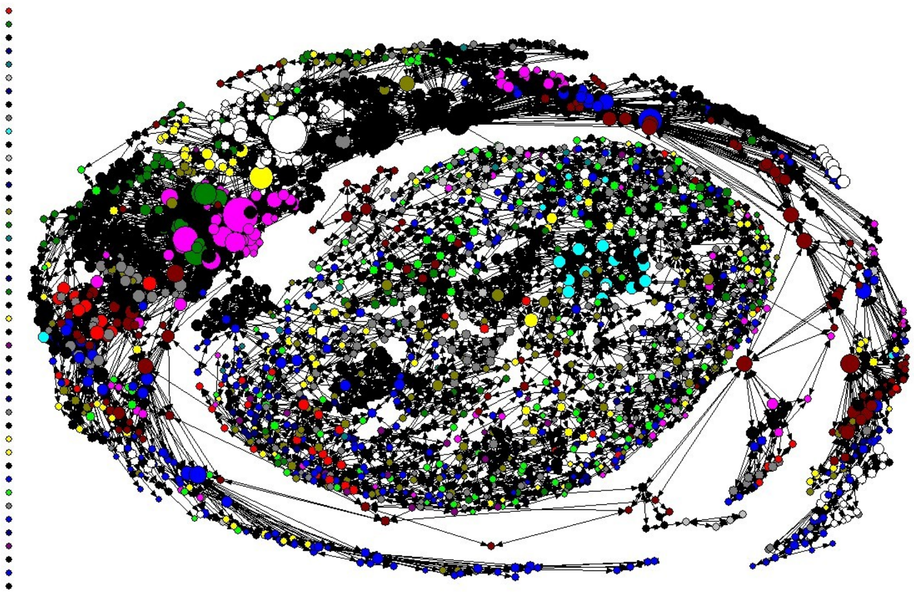

Figure 2: 2008 energy field community co-authorship social network evolution degree centrality

Figure 3 represents the same co-authorship network, where the node size represents the closeness centrality, which defines paths to the other actors. This represents the capability to monitor the information flow in the network and therefore the network activity. In this figure, we can see that the entire scientist group in the middle of the network have high closeness centrality.

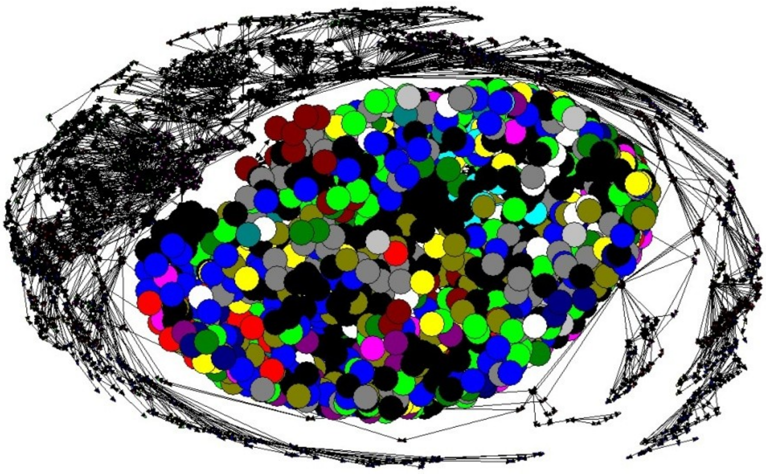

Figure 3: 2008 energy field community co-authorship social network evolution closeness centrality

Figure 4 illustrates the betweenness centrality, which is seen as the number of geodesic paths that goes through a node, expressed as a measure of centrality. This reflects the capacity of an author to connect with other authors in the network, *i.e.*, it is a measure of an author's ability to perform a "broker" role within the network (Acedo et al., 2006).

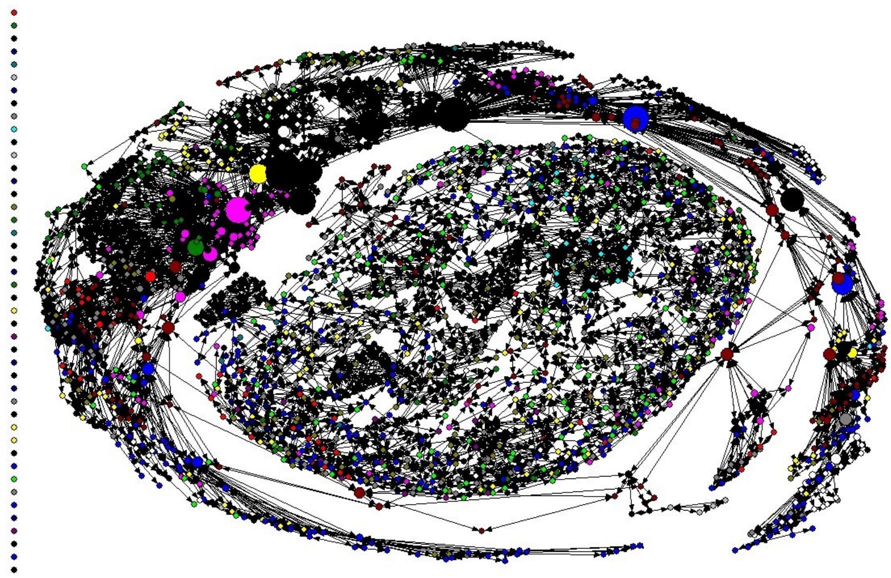

Figure 4: 2008 energy field community co-authorship social network evolution betweeness centrality

Figure 5 illustrates the eigenvector centrality, which is seen as the importance of a node in the network. An actor with a high eigenvector centrality is connected to many other actors, who themselves are well connected and are therefore the most likely to be receiving new ideas.

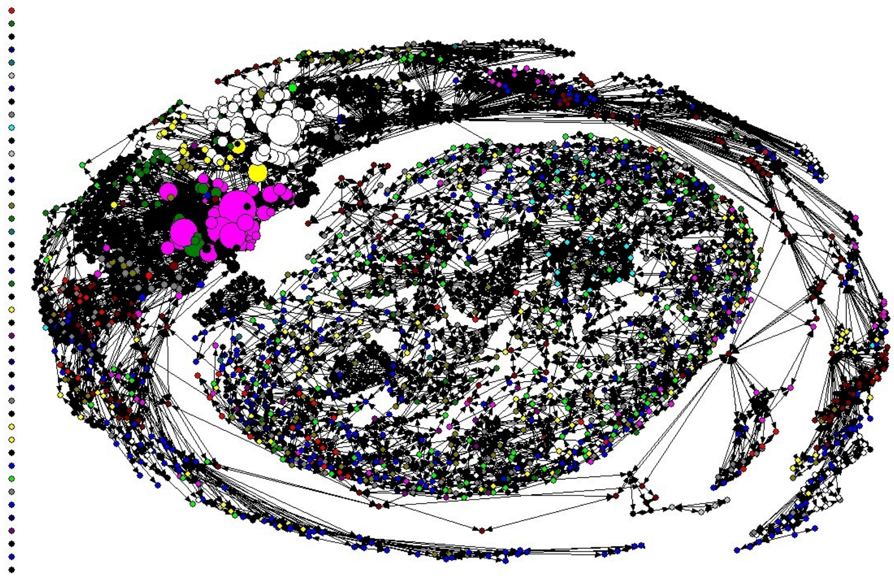

Figure 5: 2008 energy field community co-authorship social network evolution eigenvector centrality

Table VI lists the information displayed pictorially in the graphs above, in terms of the authors of the highest number of papers, degree, betweenness and eigenvector. The closeness centrality is not shown in this table, because are too many authors with the same closeness, making it difficult to choose a top 10. It is important to look at the co-authorship network as a whole and to be cautious when interpreting the predominance of authors within the network.

Table VI: The authors with highest number of papers, degree, betweeness and eigenvector

Number of papers			Degree			Betwenness $(x10^4)$			Eigenvector		
81	Gulyurtlu, I	INETI	63	Gulyurtlu, I	INETI	17	Cupido, L.	Others	0,284	Varandas, C.A.F.	EURATOM/IST
74	Cabrita, I.	INETI	62	Alves, E.	Others	16	Fonseca, R.A.	Others	0,234	Gulyurtlu, I	INETI
35	Lemos, Jose P S	UTL	60	Cabrita, I.	INETI	11	Mendonca, J.T.	UTL	0,229	Cabrita, I.	INETI
30	Pinto, F.	INETI	57	Varandas, C.A.F.	EURATOM/IST	11	Cruz, N.	Others	0,209	Malaquias, A.	Others
30	Mendonca, J.T.	UTL	41	Malaquias, A.	Others	11	Varandas, C.A.F.	EURATOM/IST	0,195	Silva, C.	EURATOM/IST
28	Varandas, C.A.F.	EURATOM/IST	38	Silva, C.	EURATOM/IST	10	Carvalho, P.	Others	0,190	Cabral, J.A.C.	EURATOM/IST
25	Abelha, P.	INETI	37	Fonseca, R.A.	Others	10	Alves, E.	Others	0,178	Fernandes, H.	EURATOM/IST
25	Carvalho, M.G.	UTL	37	Freitas, P.P.	Others	9	Cardoso, V.	Others	0,153	Boavida, D.	INETI
24	Freitas, M.C	ITN	36	Barradas, N.P.	Others	8	Lemos, Jose P S	UTL	0,151	Nedzelskiy, I.	EURATOM/IST
23	Bicudo, P.	UTL	35	Mendonca, J.T.	UTL	8	Pereira, R.	DRAM - DIAS	0,144	Pereira, L.	UTL

The scientists of the institutions *Instituto Nacional de Engenharia, Tecnologia e Informação* (INETI), *Universidade Técnica de Lisboa* (UTL) and *European Atomic Energy Community* (EURATOM) /*Instituto Superior Técnico* (IST), are very active in this energy field.

7. Further Research

The results presented in this paper are preliminary, as only around 26% of the database was used, for the reasons stated in sections 4 and 5. The next stage in this study would be to apply the same methodology to the remaining 74% of the database. The continuation of this analysis will provide a deeper knowledge as to how the network is evolving over time.

Subsequent detailed analyses will take account of political, regulatory, legal and R&D policies - at European and national level - and their combined effects on co-authorship within the energy field.

8. Conclusions

The present study is one of the first to use Social Network Analysis to examine the collaborative activities of Portuguese scientists or foreign scientists in Portugal, using the *ETDEWEB* database.

Preliminary results show that it is possible to identify important factors affecting information flows, within this network.

Using a Social Network Analysis approach to study this collaboration in the energy field, we provide a preliminary view of how this particular research community is organized and how it has evolved over time; this information is likely to be an important precursor to determining which actions are appropriate to develop this community even further, in order to increase scientific collaboration, the sharing of ideas and to improve knowledge transfer.

This preliminary study represents a starting-point which enables to understand the energy field co-authorship among the members of the community studied. It should be pointed out that because it does not address all existing data for the period 1995-2008, is important to look at the co-authorship network as a whole and to use caution when interpreting the predominant positions of the authors within the network.

The most relevant conclusion of this preliminary study is that the energy field community under study - Portuguese scientists, working either in Portugal or abroad; foreign scientists working in Portugal and scientists who have co-authored with either of these groups - has developed and increased the co-authorship relationships and there is evidence that the average number of authors per articles and conference communications has been increasing over-time. In particular, articles/papers with two and three authors represents 46,9% of the co-authorship relationships, as can be seen from tables IV and V.

The co-authorship network in the energy field seems to have two big sub-groups within, but they are connected with each other through co-authorship ties; these links could indicate that scientists belonging to different sub-disciplines of the energy field tend to collaborate. This is very interesting considering that exists in the ETDEWEB almost 50 subjects and validates the necessity to do future research considering the subjects covered.

In the data set under study a number of authors are still disconnected from the main component. This should receive also further attention, since this situation may be caused by the number of records used in this preliminary exploratory analysis not being the total universe.

References

Acedo, F. J. et al. (2006). Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis. *Journal of Management Studies*, *43*(5), 957-983.

Allen, J. et al. (2007). Formal versus informal knowledge networks in R&D: a case study using social network analysis. *R&D Management*, *37*(2), 179-196.

Barabási, A. L. et al. (2002). Evolution of the Social Network of Scientific Collaboration. *Physica A*, *311*, 590-614.

Barnet, A. H. et al (1988). The Rising Incidence of co-authorship in economics: further evidence. *Review of Economics and Statistics*, *70*(3), 539-543

Borgatti, S. P. (2002). *Netdraw: Graph Visualization Software*. Harvard Analytic Technologies [http://www.analytictech.com/Netdraw/netdraw.htm] last accessed 29 March 2009

Clark, L. (2006). *Network Mapping as a Diagnostic Tool*. Centro Internacional de Agricultura Tropical, La Paz, Bolivia.

Cronin, B. (1996). Rates of return of citation. *Journal of Documentation*, *52*(2), 188-197.

Hara, N. et al. (2003). An Emerging View of Scientific Collaboration: Scientists' Perspectives on Collaboration and Factors that Impact Collaboration. *Journal of the American Society for Information Science and Technology*, *54*(10), 952-965.

Katz, J. & Martin, B. R. (1997). What is research collaboration?. *Research Policy*, *26*, 1-18.

Kilduff, M. & Tasi, W. (2003). *Social Networks and Organizations*. London: SAGE Publications.

Laband, D. N. & Tollison, R. D. (2000). Intellectual Collaboration. *Journal of Political Economy*, *108*(3), 632-662.

Li-chun, Yin et al. (2006). Connection and Stratification in research collaboration: An analysis of the COLLNET network. *Information Processing and Management*, 42, 1599-1613.

Moody, J. (2004). The Structure of a social science collaboration network: disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69, 213-238.

Neto, M.C. et al. (2008). Social Network Analysis Applied to Knowledge Creation and Transfer in the Portuguese Agricultural R&D Field: an exploratory study. Harorima, D. & Watkins, D. (2009). *Proceedings of the 9th European Conference on Knowledge Management*. Southampton, UK. ISBN: 978-1-906638-10-8

Newman, M. E. J. (2001). Scientific collaboration networks. *Physical ReviewE*, *64*.

Piette, M. J. and Ross, K. L. (1992). An analysis of determinants of co-authorship in economics, *Journal of Economic Education*, 23, 277-283

Vidgen, R. et al. (2007). What sort of community is the European Conference on Information Systems? A social network analysis 1993-2005. *European Journal of Information Systems*, *16*, 5-19.