2015 International Nuclear Atlantic Conference - INAC 2015
São Paulo, SP, Brazil, October 4-9, 2015
Associação Brasileira de Energia Nuclear - ABEN
ISBN: **978-85-99141-06-9**

COLLABORATION NETWORKS AND RESEARCH PRODUCTIVITY AT IPEN

Carlos Anisio Monteiro and Antonio Carlos de Oliveira Barroso

Instituto de Pesquisas Energéticas e Nucleares (IPEN / CNEN - SP)
Av. Professor Lineu Prestes 2242
05508-000 São Paulo, SP
monteiro@ipen.br, barroso@ipen.br

ABSTRACT

In this article, we investigate the IPEN’s scientific collaboration network. Based on publications registered in IPEN's technical and scientific database was extracted a set of authors that developed technical and scientific work on the 2001 to 2010 period, using coauthorship to define the relationship between authors.
From the data collected, we used degree centrality indicator in conjunction with two approaches to assess the relationship between collaboration and productivity: normal count, where for each publication that the author appears is added one for the author’s productivity indicator, and fractional count which is added a fractional value according to the total number of publication's authors. We concluded that collaboration for the development of a technical and scientific work has a positive correlation with the researchers productivity, that is, the greater the collaboration greater the productivity.
We presented, also, a statistical summary to reveal the total number of publications and the number of IPEN's authors by publication, the average number of IPEN's authors per publication and the average number of publications by IPEN's author, the number of IPEN's authors that not published with no other author of the IPEN and, finally, the number of active and inactive (ex. retirees) researchers of the IPEN, as well as, the number of authors who do not have employment contract with the IPEN.

1. INTRODUCTION

It is well known and several studies show the importance of collaboration for the generation, preservation and dissemination of knowledge and therefore as a way to boost the productivity see [1][2][3][4][5].

Collaboration is a reality that is present in the daily life of scientific and technological communities. Also in last decades this subject has evoked great interest in the business community, because many innovations and new technological developments have been achieved through collaborative efforts. Advances of information and communication technology have enabled new forms of collaboration and also new tools became available to study in more detail the characteristics and consequences of collaborative work.

In R&D environments, the increasingly interdisciplinary research, the complexity of the scientific problems, and the rising cost of equipment and other scientific resources that are characteristic of modern science encourage researchers to establish scientific collaboration links [1][6].

We may say that scientific collaboration is as a socio-technical process that occurs naturally and seems to be a good practice to leverage knowledge creation, transfer and preservation, as well as technological development. The source of virtuousness of the whole process is well described in the words of the famous writer George Bernard Shaw.

> *"If you have an apple and I have an apple and we exchange these* apples *then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas." George Bernard Shaw.*

However, there is no clear definition of the term scientific collaboration [7], but there is a consensus that it is a process of interaction between two or more researchers to achieve a common goal. For example, [8] defined collaboration as the interaction that occurs within a social context between two or more researchers that facilitates the sharing of meaning and performing tasks in relation to a shared goal and, [9] defined collaboration as the joint efforts of researchers to achieve a common goal of producing new knowledge. [10] point out that collaboration is a social process that involves human interaction and can occur in various ways and for different reasons.

Scientific collaboration is a topic that is studied in various disciplines including information science, psychology, management science, sociology, computer science, philosophy, social studies and any other discipline in which scientific collaboration may be present [8].

There are several reasons why researchers collaborate in the development of an R&D project, as can be seen in Table 1.

Table 1: The reasons why the researchers collaborate.

a) Acess to expertise;
b) To have access to equipments, resources and funding for research projects;
c) To obtain prestige and visibility to advance professionally and progress more rapidly;
d) To enhance productivity;
e) To expand the networking;
f) To learn new skills or techniques;
g) To satisfy the curiosity or intellectual interests;
h) to share the excitement of an area with other people;
i) To reduce errors;
j) to reduce isolation, recharging energy and the excitement;
k) To promote academic support for students;
l) To enhance knowledge and learning.

Adapted from [11].

In the case of IPEN, an R&D institution that can be viewed as a system for the production of knowledge and technology, ensuring the generation, preservation and dissemination of knowledge is an essential condition for the survival and development of the institute. In this context, understanding how its researchers collaborate is of crucial importance, especially, to preserve nuclear knowledge and expertise that have been accumulated over decades [12]. To get an idea of the knowledge generated at IPEN from 2001 to 2010 period, see [13].

From IPEN's technical and scientific database, a subset of it containing the relevant information pertaining to all publications and authors during the period 2001-2010 was extracted, using coauthorship to define the relationship between authors. From this relationship data, some hypotheses were made trying to related the networking capability and position of an author to its productivity. To assess these hypotheses degree centrality indicator arising from social network analysis (SNA) technique [14][15] were used in two different approaches that will be fully discussed in the text.

2. DATA AND METHODOLOGY

Data collection was based on coauthorship, from 2001 to 2010 period. A computer program, developed internally, extracted from the IPEN´s technical and scientific publications database, the authors and coauthors of IPEN, as well as the amount of publications produced in coauthorship. The database compiles records of dissertations, theses, articles in periodics, books, book chapters and articles from conferences and other scientific events.

The computer program generates a symmetric adjacency matrix ($n_{ij} = n_{ji}$), where an entry n_{kl} indicates the number of publications coauthored by persons k and l during a period of two years. Time slices of two years were considered, thus generating data for five networks. These networks can be viewed as longitudinal photos of a dynamic network. Authors were identified by the last name and the initial letters of their first names, since there is no a related identification code for each author. It is important to note that this fact may have caused an underestimation or overestimation of the overall number of authors and of publications related to these authors, as this may have generated an ambiguity in the interpretation of these metrics.

The productivity of each author of IPEN was measured based on two approaches quoted by [16]:

a) normal count (*NC*): where it is computed an unit for each author involved in the publication. For example, in a publication with three coauthors is computed one publication for each of the three coauthors.

$$NC_j = N_j \qquad (1)$$

Where *Nj* represents the total number of publications in which the *j* researcher appears as author or coauthor. This metric is a bit inflationary, as the total number of publications does not matches the *Nj* summation.

b) fractional count (*FC*): is computed one fraction for each author and coauthor involved in the publication. For example, in a publication with one author and three coauthors are computed 1/4 of unit for each of the four coauthors.

$$FC_j = \sum_{i=1}^{N_j} \frac{1}{n_i} \qquad (2)$$

Where n_i is the total number of authors of the *i-th* publication in that *j* researcher appears. This type of approach eliminates the inflation and by using both one can get interesting conclusions.

The number of coauthors of a specific author, that is, how many coauthors a specific author collaborated to produce its publications in each two-year period, coincides with a network indicator named (unscaled) degree centrality (*DC*) of the dichotomized matrix obtained from the adjacency matrix referred in the beginning of the section.

3. RESULTS AND DISCUSSION

3.1. Bibliometrics Data

In the Table 2 we provide a summary of the technical and scientific production of IPEN from 2001 to 2010, grouped in consecutive two years intervals.

Table 2: Summary of IPEN´s technical and scientific production.

corresponding year for the period	2002	2004	2006	2008	2010
Periods	*2001 2002*	*2003 2004*	*2005 2006*	*2007 2008*	*2009 2010*
Total of publications	1,559	1,261	1,557	2,042	2,042
Validated Publications	1,456	1,191	1,459	1,946	1,972
Publications with IPEN´s authors not identified	103	70	98	96	70
IPEN´s actives	360	287	327	344	351
Autores inativos (aposentados)	42	35	41	33	46
Others outside IPEN	1,444	1,341	1,999	2,389	2,379
IPEN’s authors per publication	1.983	1.865	1.935	1.851	1.795
Total authors per publication	3.932	3.860	4.163	4.219	4.188
Publications per IPEN’s authors	7.182	6.898	7.671	9.557	8.914
Publications with one IPEN’s author	665	565	713	996	1.128
Publications with two IPEN’s authors	433	374	408	514	431
Publications with three IPEN’s authors	199	162	191	274	221
Publications with four IPEN’s authors	85	50	76	94	125
Publications with five IPEN’s authors	45	28	38	37	39
Publications with six IPEN’s authors	18	8	15	23	14
Publications with more than six IPEN’s authors	11	4	18	8	14
IPEN's authors that does not published any other IPEN's author	16	21	19	9	13
Total of thesis and dissertations	261	224	264	268	311

A careful look into the data from Table 2 reveals that production of publications has increased thanks to a productivity increase of the IPEN's authors and also through an increase of collaboration with outside authors. From the first to the second biennium there was a decrease in productivity of IPEN's authors very related to the decrease of external collaboration. This was followed by two periods of a vigorous productivity increase, which seems to be supported by the increase of external and internal collaboration. It is interesting to note that only in the first of these biennia (2004-2005) some rise of internal collaboration was also observed (IPEN´s authors per publication line in Fig. 1). In the last biennium, the average number of authors per publication (Total authors per publication line) stayed leveled and IPEN's authors per publication have continued to decrease, this indicates a compensating growth in the number of external collaborating authors. It should be noted that there was a drop in the productivity of IPEN's authors in this last period.

It must be noted that it is being denoted as external coauthors people from other institutions, but also graduate students engaged in master and doctoral research at IPEN-USP Post-Graduation Program and those account a significant portion of the external coauthors.

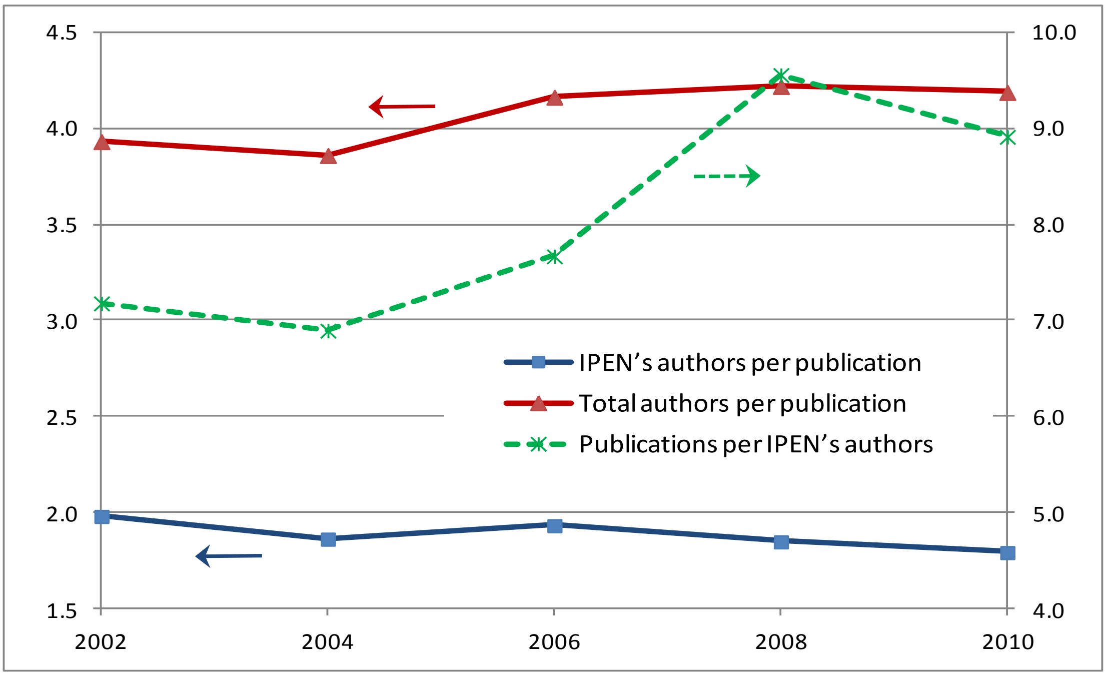

Figure 1: Productivity and collaboration.

The analysis of the overall tendencies shows that from 2001 to 2010 there was a sound 24.1% increase in the publication productivity of IPEN's authors. In the same period IPEN's author population has decreased by 1.2%, but the number of external collaborators has increased by 64.7%. The population of external collaborators has leveled out in the last 3 years and since a significant portion of them are students, this is an indication that IPEN's PG has reached a plateau. This can be seen in Fig. 3, that also shows the authors population in IPEN's scientific collaboration network and the total number of publications that had at least one IPEN's author identified (validated publications).

From the previous analysis it can be concluded that fostering more and diversified collaboration among researchers and also to have a large population of graduate students helps knowledge and technology production at IPEN. From an overall perspective these two factors should deserve a lot of attention and support from IPEN's managers.

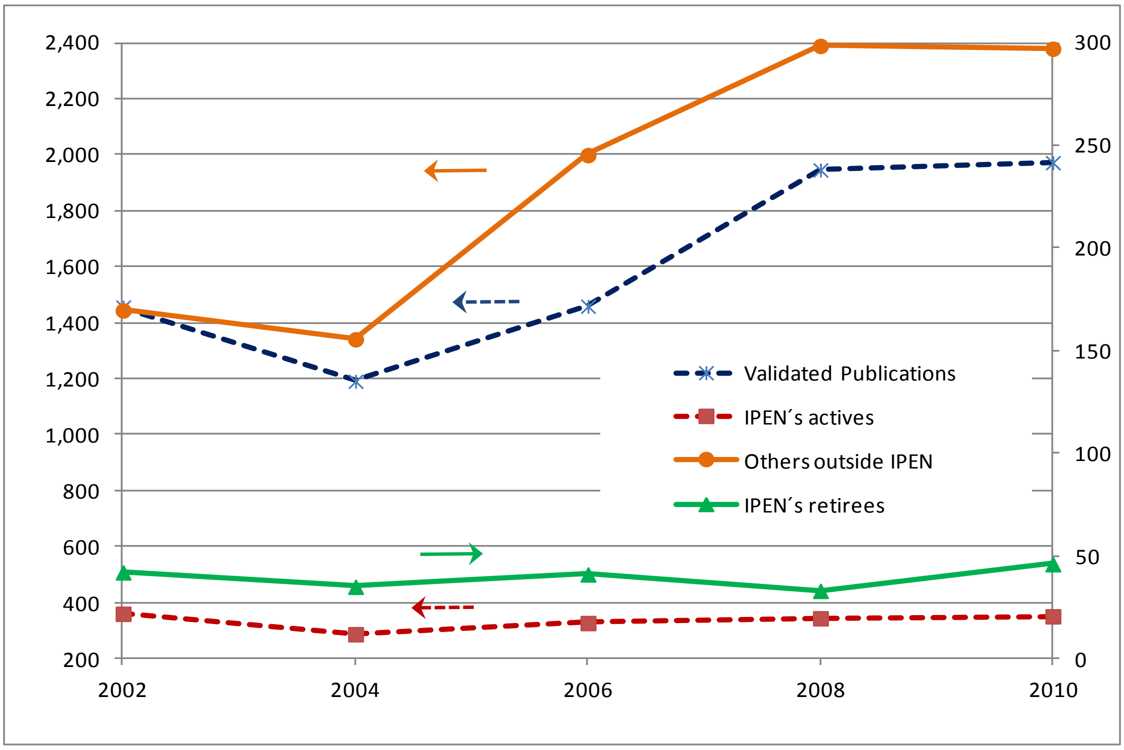

Figure 2: Publications and authors population.

3.2. Productivity Indicators

The *NC* counter corresponds to an inflated value, since for each author and coauthor of IPEN is added one publication. As a result, the counter that best depict the productivity of IPEN's authors is the *FC*, where for the productivity rates of each author or coauthor is stipulated a fraction which is calculated based on the total number of published authors, regardless of whether these authors and coauthors are from the IPEN or not. Thus, using as a basis the *FC* counter, the productivity of the authors and coauthors of IPEN referring to technical scientific publications corresponding to approximately 57%, 53%, 52%, 49% and 47% for the periods 2001/2002, 2003/2004, 2005/2006, 2007/2008 and 2009/2010, respectively. These percentages can be abstracted from Fig. 3.

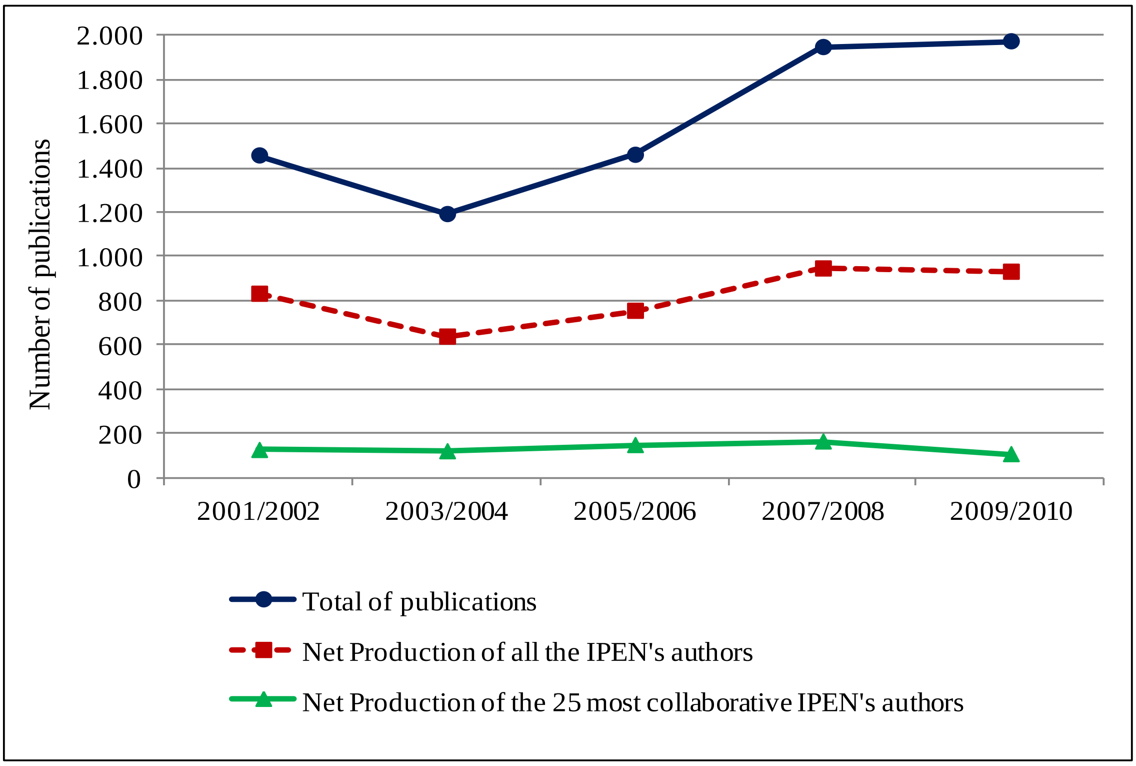

Figure 3: Net production of IPEN's authors.

The generation of a publication with a larger number of IPEN's authors or coauthors represented by degree centrality indicator (*DC*) does not necessarily indicate a higher productivity. As an example, compare in Table 3, the number of coauthors (*DC*) and the amount of publications (*NC*) of *A127* author with that of the *A251* author and, we can observe that the development of R&D work with a greater number of IPEN's coauthors not reflected in a greater number publications in all periods, except the 2009/2010 period when the *A251* author does not appear between the 25 more collaborative authors.

Table 3: Productivity of the 25 IPEN's authors that more collaborated with other IPEN's authors.

2001/2002				2003/2004				2005/2006				2007/2008				2009/2010			
Author	DC	NC	FC	Author	DC	NC	FC	Author	DC	NC	FC	Author	DC	NC	FC	Author	DC	NC	FC
A284	22	57	15.2	A251	24	62	14.2	A298	23	9	1.2	A127	28	38	9.1	A330	24	10	2.1
A021	21	20	4.9	A360	24	8	1.4	A127	20	34	8.2	A105	22	15	2.7	A296	23	7	1.3
A360	21	26	4.4	A127	21	19	4.7	A178	20	18	3.2	A271	22	14	2.2	A127	22	17	3.5
A127	20	27	6.0	A284	21	48	11.2	A018	19	13	2.4	A298	20	15	2.5	A098	21	77	18.1
A025	19	20	4.7	A308	19	6	1.2	A227	19	6	0.7	A018	19	10	1.8	A105	21	20	3.0
A220	19	12	2.4	A349	19	11	2.1	A105	18	5	0.7	A314	18	38	10.9	A271	21	21	4.1
A442	19	11	2.7	A196	16	9	1.8	A015	17	26	4.7	A098	17	112	25.3	A068	19	11	1.9
A330	18	15	2.4	A330	16	8	1.7	A098	16	72	14.8	A251	17	51	12.2	A321	19	4	0.5
A011	17	28	8.4	A041	15	22	5.3	A236	16	41	7.8	A330	17	16	4.7	A041	18	30	6.8
A201	17	5	0.8	A242	15	28	6.7	A284	16	68	19.8	A011	16	10	3.2	A172	18	5	1.1
A255	16	21	12.3	A001	14	16	3.4	A359	16	9	1.8	A022	16	38	8.0	A219	18	60	12.7
A274	16	39	11.2	A009	14	13	3.0	A025	15	6	1.2	A236	16	49	9.8	A236	18	36	6.6
A109	15	21	8.1	A048	14	18	4.2	A242	15	42	10.1	A284	16	58	15.9	A314	18	21	5.7
A236	15	26	5.7	A077	14	18	3.9	A341	15	31	7.3	A042	15	59	13.4	A298	17	23	3.9
A251	15	46	11.3	A145	14	5	1.6	A360	15	12	3.1	A219	15	51	11.7	A337	17	19	4.3
A013	14	21	3.5	A182	14	17	3.4	A154	14	5	0.7	A216	14	11	2.8	A002	16	11	2.3
A093	14	16	4.5	A274	14	47	14.5	A220	14	13	2.7	A025	13	12	2.9	A064	16	19	5.9
A322	14	7	1.8	A472	14	2	0.2	A251	14	44	10.7	A027	13	21	7.0	A074	16	9	1.3
A349	14	11	2.2	A118	13	14	2.9	A299	14	19	4.4	A259	13	15	3.6	A118	16	6	1.5
A497	14	4	0.5	A216	13	4	0.9	A304	14	34	8.2	A307	13	9	1.4	A238	16	24	5.0
A042	13	19	3.6	A236	13	15	2.9	A352	14	18	5.8	A048	12	9	2.0	A011	15	9	2.3
A118	13	22	3.6	A283	13	23	7.9	A011	13	13	4.1	A056	12	23	4.0	A113	15	4	0.5
A143	13	6	1.4	A341	13	16	3.1	A041	13	28	6.6	A058	12	5	0.8	A336	15	26	4.6
A162	13	5	1.0	A362	13	50	13.3	A132	13	57	12.0	A160	12	16	4.7	A348	15	15	4.9
A216	13	10	1.9	A011	12	13	3.6	A160	13	16	4.6	A191	12	3	0.4	A360	15	10	1.9

Note: ordered by DC indicator.

However, in order to evaluate how this aspect related to productivity behaved in relation to all authors and coauthors in each network, we calculate the correlations based on *Pearson's* coefficient, between the *DC* indicator obtained from dichotomized networks and the indicator *NC* productivity of all authors and coauthors for each biennial network. The *Pearson's* coefficient [17][18] is a measure of the linear correlation between two variables. The values range from -1 to +1, which indicates the intensity and the direction of the association. The direction of the association when the coefficient is positive indicates that changes in variables values tend to follow in the same direction, while a negative coefficient indicates that the variables values are going opposite directions. Table 4 presents some practical rules that are adopted to evaluate the intensity of the association.

Table 4: Practical rules to evaluate the strength of the correlation coefficient.

Coeficient´s variation	*Correlation intensity*
± 0.91 - ± 1.00	Very high
± 0.71 - ± 0.90	High
± 0.41 - ± 0.70	Moderate
± 0.21 - ± 0.40	Low, but defined
± 0.01 - ± 0.20	Negligible

Source: [18, p. 312].

The values of the correlation and statistical significance tests were calculated using the *cor.test()* function available in the *R* statistical program and, specifying the "*pearson*" how extraction method, which is based on *Pearson's* correlation coefficient. An example of results obtained with *cor.test()* function is shown below:

```
> cor.test(DC0910,NC0910,method="pearson",alternative="two.sided")
Pearson's product-moment correlation
data:  DC0910 and NC0910
t = 10.1102, df = 395, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.3715642 0.5282537
sample estimates:
cor
0.4534051
```

```
> cor.test(DC0910,FC0910,method="pearson",alternative="two.sided")
Pearson's product-moment correlation
data:  DC0910 and FC0910
t = 7.0405, df = 395, p-value = 8.542e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.2434918 0.4185772
sample estimates:
cor
0.3339115
```

How it is observed in Table 5 there is a positive correlation with moderate intensity between the number of IPEN's coauthors and the productivity, noting that the development work in collaboration with a larger number of IPEN's authors tends to increases productivity, as opposed to what was previously observed when the analysis was restricted to 25 more collaborative authors.

Table 5: Values of the coefficients calculated for the correlation between the *DC* indicators and the *NC* indicator.

Network (period)	Degree Centrality	
	Normal count	Fractional count
2001/2002	0.569	0.411
2003/2004	0.557	0.409
2005/2006	0.534	0.404
2007/2008	0.529	0.446
2009/2010	0.453	0.334

Note: All correlations are significant at $p < 0.001$

4. CONCLUSIONS

In this article, we showed that there is a positive correlation with moderate intensity between collaboration and productivity related with the authors who make up the R&D network of IPEN.

As this correlation was determined only in relation to researchers from IPEN, we show that there is a very strong indication that the collaboration between IPEN’s researchers can result in greater productivity, both in the number of real publications, where equal credit is given to all authors, as in a number of publications weighted by number of coauthors.

It also showed that there was a significant increase in the number of external authors to IPEN that are contributing to increased productivity. However, he warned that if the IPEN not adopt policies to retain this manpower, the process of preservation of knowledge is lame, endangering the future of research in IPEN.

Finally, we present a summary of bibliometrics data associated with the R&D network of IPEN.

REFERENCES

1. S. Lee, B. Bozeman, "The Impact of Research Collaboration on Scientific Productivity," *Social Studies of Science*, 35, pp. 673-702 (2005).

2. A.C.O. Barroso, J.S.B. Reis-Junior, C.A. Monteiro, A. Seary, “Preservation of knowledge through networking with retirees,” *Procedia – Social and Behavioral Sciences*, **2**, pp. 6475-6482 (2010).
3. F.S. Parreiras, A.B.O. Silva, R.F. Matheus, W.C. Brandão, "RedeCI: colaboração e produção científica em ciência da informação no Brasil," *Perspectivas em Ciência da Informação*, **11**, pp. 302-317 (2006).
4. S.E. Hampton, J.N. Parker, "Collaboration and Productivity in Scientific Synthesis," *BioScience*, **61**, pp. 900-910 (2011).
5. E.F.T. Oliveira, M.C.C. Grácio, "Scientific Collaboration Network in “Metrical Studies”: A Co-Authorship Study Using The Scielo Information Science Periodicals," *Brazilian Journal of Information Science*, **2**, pp. 33-47 (2008).
6. R.B. Duque, M. Ynalvez, R. Sooryamoorthy, P. Mbatia, D.B.S. Dzorgbo, W. Shrum, "Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research in Developing Areas," *Social Studies of Science*, **35**, pp. 755-785 (2005).
7. C. Hu, P. Racherla, “Visual representation of knowledge networks: A social network analysis of hospitality research domain,” *International Journal of Hospitality Management*, **27**, pp. 302-312 (2008).
8. D.H. Sonnenwald, “Scientific collaboration,” *Annual Review of Information Science and Technology*, **41**, pp. 643-681 (2007).
9. J.S. Katz, B.R. Martin, “Distribution of Scientific Productivity: Ambiguities in the Assignment of Author Rank,” *Scientometrics*, **20**, pp. 131-144 (1991).
10. S.A.S. Vanz, I.R.C. Stumpf, “Colaboração científica: revisão teórico-conceitual,” *Perspectivas em Ciência da Informação*, **15**, pp. 42-55 (2010).
11. D.D. Beaver, “Reflections on scientific collaboration (and its study): past, present, and future,” *Scientometrics*, **52**, pp. 365-377 (2001).
12. M.E.D. Acar, A.C.O. Barroso, “A modern framework for knowledge preservation at government R&D organizations,” *Proceeding of Meeting*, International Nuclear Atlantic Conference, Recife, PE, Brazil, November 24-29, 2013, p. 10, (2013).
13. C.A. Monteiro, A.C.O. Barroso, “How Knowledge and Technology are Created in a Research Institute,” *Proceeding of Meeting*, International Nuclear Atlantic Conference, Belo Horizonte, MG, Brazil, October 24-28, 2011, p. 14, (2011).
14. S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, USA (1995).
15. J. Scott, *Social Network Analysis – A Handbook (ed. 2)*, Sage Publications, London, England (2000).
16. N. Pravdic, V. Oluic-Vukovic, “What is research collaboration?,” *Research Policy*, **26**, pp. 1-18 (1997).
17. J.F. Hair Jr., B. Babin, A.H. Money, P. Samuel, *Fundamentos de métodos de pesquisa em administração*, tradução: Lene Belon Ribeiro, Bookman, Porto Alegre, Brasil (2005).
18. J.F. Hair Jr., W.C. Black, B.J. Babin, R.E. Anderson, *Multivariate Data Analysis (ed. 7)*, Pearce Prentice Hall Inc., Upper Saddle River, USA (2010).