

# HOW KNOWLEDGE AND TECHNOLOGY ARE CREATED IN A RESEARCH INSTITUTE

**Carlos Anísio 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

Collaboration within and across R&D teams is a naturally occurring socio-technical process and it seems to be a very good thrust for technology development and knowledge creation, transfer and preservation in R&D organizations.

The study has focused on evidence gathered from: (a) the publication data base of IPEN from 2001 to 2008 was used to generate four time evolving networks of co-authorship, considering time slices of two years; (b) authors "demographic" data were included as node attributes in the networks; and (c) questionnaire surveys and interviews to understand the reasons which influence a person to search and choose partner to co-work in a research.

Social network analysis was used extensively to understand the network configurations, their time evolution, the key actors, groups and their role in this network. A few indexes and algorithms were used to get insight into the networks structures. A variety of centrality indicators were used to better characterize key actors and better understand their position and role in the network. Also many grouping techniques were used to find the most prominent/active groups.

For the most central authors or groups, demographic data was cross analyzed with their network indexes to get check for a few hypothesis of preferential attachment. Finally, based on the proper theoretical background and the information gathered with questionnaires and interviews a model was devised to explain the propensity to collaborate.

This paper reports the general ideas of the whole research and presents the most important results of the network analysis.

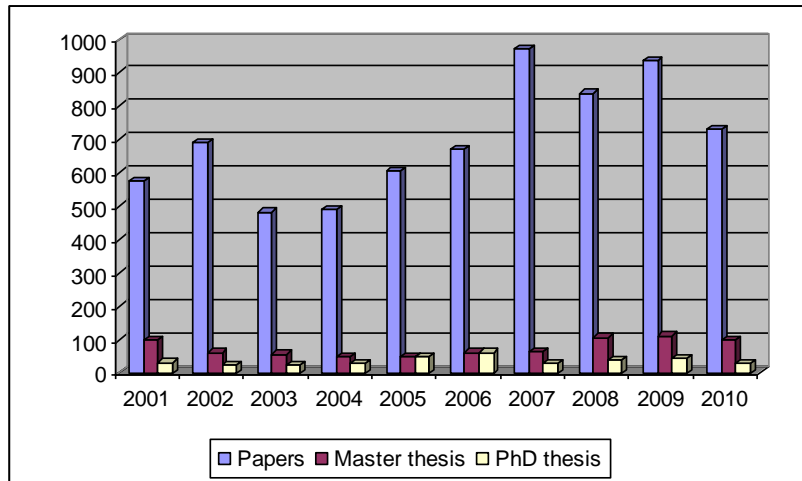
## 1. INTRODUCTION

IPEN is a research and development institution (R&D) that was traditionally focused in the areas of nuclear power and applications of nuclear techniques. During the 90's, with the deceleration of the Brazilian nuclear program, IPEN has been directed to expand and diversify its areas of interest, seeking new lines of research on topics such as biotechnology, fuel cells, laser technology, nuclear physics, radiochemistry, materials, environment and management.

As a result of its R&D activity, IPEN has developed production capacity and has begun to supply certain products and services to society. In this respect the radiopharmaceutical production unit that supplies more than 90% of national demand is very emblematic.

As a consequence of its diversified fields of research, IPEN is also a large generator and disseminator of scientific and technological knowledge. All this knowledge can be evidenced by the number of reports and scientific articles that are annually published in national and

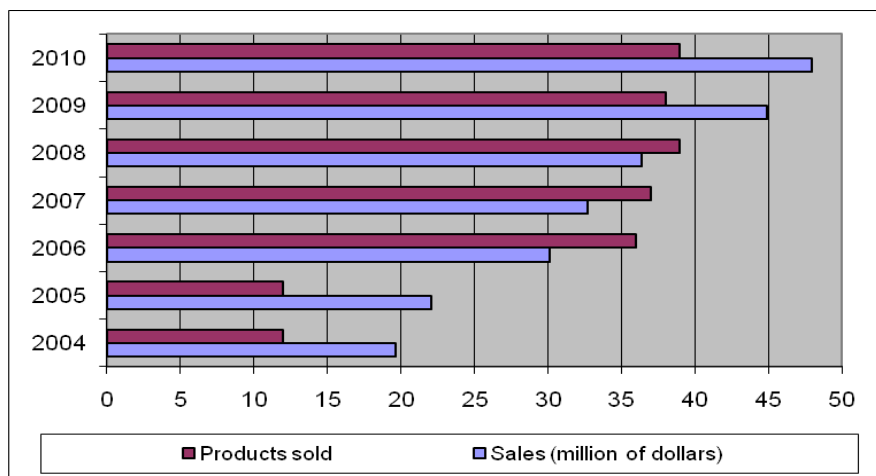
international magazines and conferences. This facet of IPEN is leveraged by its postgraduation program in nuclear technology in association with the University of São Paulo (USP), which has been ranked in the last seven years as the best in the country. Fig. 1 shows the number of scientific articles, master thesis and PhD thesis produced in the period 2001 to 2010.



**Figure 1. IPEN's Technical-Scientific Production in the period 2001 to 2010**

*(Source: IPEN-PTC database)*

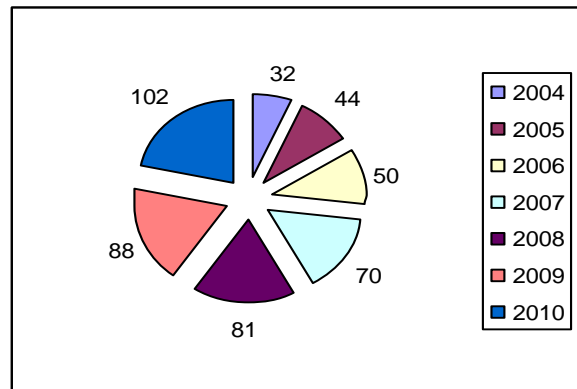
To give an idea of the magnitude of IPEN's activities supplying products and services, an overview of the evolution of sales and number of products sold by the institution is shown in Fig. 2.



**Figure 2. Radiopharmaceuticals production and sales**

*(Source: IPEN SIGEPI)*

Another factor that puts the institution as protagonist in the generation of knowledge is its search for innovation. This is reflected in the increasing number of patent applications that are deposited in the last years, as shown in Fig. 3.



**Figure 3. Number of patents developed at IPEN**

*(Source: IPEN SIGEPI)*

IPEN's performance, in its multiple activities, is fundamentally related to the quality and quantity of knowledge that the institution produces. Focusing on how this knowledge is produced, IPEN can be seen as a conglomerate of scientific and technological collaboration networks, probably the majority of them informal networks, covering hundreds of people from different units of IPEN and also from other R&D institutions. There are several factors that leverage the collaboration between researchers in R&D environments, as described from different perspectives by [1,3,15].

This paper presents a theoretical framework and subjacent methodology for the purpose of understanding the phenomena of knowledge creation, preservation and transfer in R&D institutions, and especially IPEN. For this purpose, two methodological approaches were employed:

- i. From the perspective of the social network analysis (SNA) [13,17,21], it was used the methodology and tools to map evidence based networks and to evaluate various indicators that identify the roles and prominence of people and groups in this IPEN's scientific collaboration network in R&D;
- ii. Structural equation modeling (SEM) techniques were used to study which personal and organizational attributes and/or characteristics, influence people's propensity to establish collaboration in R&D environments.

Within the so called knowledge society [1,9,10,14,16,18,19], knowledge or intellectual capital becomes the great engine for sustainable development and success of an organization. However knowledge cannot be crystallized, in fact must be continually shared between people and applied to creating / improving technology, products, services and new knowledge. Therefore, this alluded process of creative interaction is somehow a (dynamic) photo of the organization's social capital. The concept just described somehow resonates and adds specificity to those presented by [6,7] that defines social capital as an intangible asset

that is embedded in relationships between people and [8] who stated that it can be understood as the people you relate whom have a significant impact on your knowledge. In short it can be stated that this work presents a complete method to assess those dynamic photos of the organization's social capital as well as its longitudinal (time) evolution.

## **2. HOW DOES THIS HAPPEN?**

To examine this process, several research questions were raised:

- (a) Do researchers work isolated?
- (b) Do they prefer to collaborate in small or large groups?
- (c) Are the groups separated or share many elements?
- (d) What does set these preferences?
- (e) What evidences can be taken to understand these preferences?
- (f) What personal and organizational characteristics influence on people's propensity to establish research collaboration?
- (g) Is this collaborative framework related to productivity?

These questions suggested four major research objectives, which are:

- (a) To understand how knowledge is created, transferred and preserved in IPEN;
- (b) To analyze what characterizes the most prolific and collaborative researchers, more specifically, who are those with large number of collaborating partners and production evidence (papers) and what influences this dynamic;
- (c) Examine what makes a given researcher more attractive for the others to establish collaborative links with him / her;
- (d) Provide IPEN and consequently the CNEN, a theoretical and methodological framework, based on evidence and modeling, to assist in the implementation of policies and initiatives of knowledge creation, preservation and transfer in their units.

Other objectives emerge naturally due to the variety and analytical capacity of the SNA methodology. An incomplete list follows:

- (a) To reveal insights from IPEN's R&D collaboration network configuration;
- (b) To disclose who collaborates with whom and what are most relevant underlying causes for such;
- (c) To identify who are the most productive researchers and research teams;
- (d) To identify how per capita productivity of research groups varies as a function of size, density and cohesion of their connections;
- (e) To reveal if and how the structure of the researchers' network and subnets have similarity with the formal organizational structure and to find possible explanations.

In relation to the MEE approach some specific objectives are partially listed below:

- (f) Based on a statistically qualified dataset, to optimize the model, while still respecting the criteria of parsimony;
- (g) Using a test dataset to validate the optimal model and possibly a variant of it.

Finally from the combination of the two approaches two higher-level goals emerged as follows:

- (h) On the one hand, determine to what extent the propensity to cooperate can explain, even partially, the network structure and/or some key indicators achieved from the ARS approach;
- (i) On the other hand, understand the extent to which the structural position of an individuals or groups on the network amplifies how they are perceived by the others making them more desirable to collaborate with.

### **3. FORMULATING THE PROBLEM APPROACH**

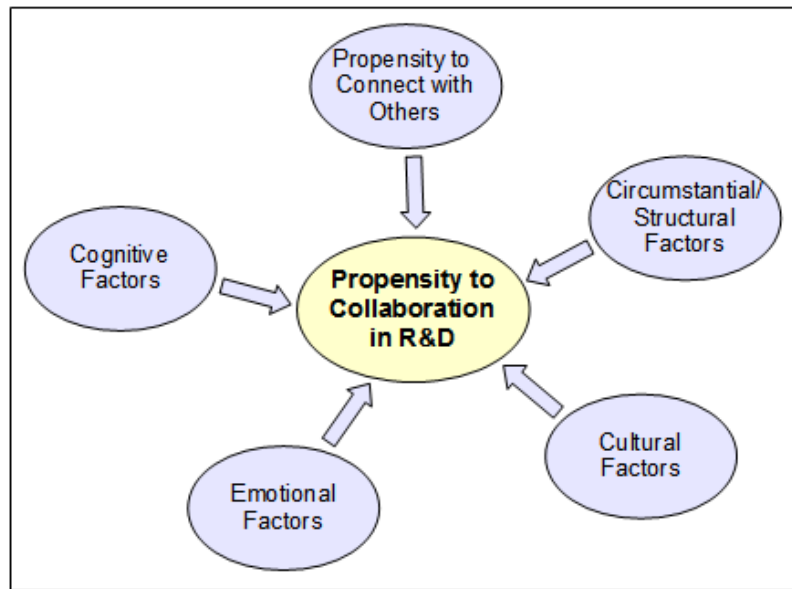
The first approach uses the SNA paradigm and based on the publication and human resources databases of IPEN the R&D networks were assembled. For each network several centrality indicators (degree, betweenness, boundary spanners, eigenvector, hub, etc.) were calculated to provide a better characterization of the key actors and a better understanding their position and role in the network. At the group level, many techniques were used to disclose the most prominent and active groups in the network by inspecting, for example, cliques, plexes and Newman and Concor groups. Finally, at the network level a range of indicators related to density was analyzed, such as: fragmentation, cohesion, robustness, clustering coefficient, average distance, effective path, etc., to explain the network structure, configuration and topology, as well as, its temporal evolution in the eight years of the study.

The main elements, for which data is gathered and upon which all the social network analyses are done, are cleared and/or defined as follows:

- The system - IPEN or more specifically all the actors/agents involved in the creation/development of knowledge/technology;
- The networks – defined by collaborative relationships between the actors and groups and that result in the knowledge/technology production;
- The artifacts of this production (hard data) - characterized by technical-scientific articles and reports;
- The actors' objective attributes – defined as his hierarchical position, history of work at IPEN, grants, productivity fellowship, formal education background, etc.

In the second approach, more subjective variables are taken into account. Personal attributes and/or characteristics are linked to cognitive, emotional and cultural factors. While the organizational aspects will be set by structural and circumstantial factors related to research centers and/or areas of the institute. The propensity for collaboration in R&D, in this aspect, is the main construct or latent variable and will be evaluated by other latent and observable variables based on aspects related to personal and organizational factors mentioned.

The proposed model to explain what factors have greater influence on the population of the knowledge actors at the time of choosing a collaborating partner is depicted in Fig. 4. By knowledge actors, or just actors, is understood all IPEN's researchers, technologists, analysts and technicians - including retirees - that through their collaborating interactions create or enhance knowledge, methodologies, processes, as well as develop new products and services.



**Figure 4. Theoretical model**

In addition to the factors shown above, the proposed model will include an adaptation of the model developed by [20], which examined individual differences that influence the propensity of people to connect with other people in different environments - academic and organizational. This model includes three latent variables, "make friendships", "make acquaintances" and "join people".

#### **4. METHODOLOGY**

From the publication database, considering the period 2001 to 2010, it was computed co-authorship relationships matrix, grouped in periods of two years. For each pair of effective collaborating actors the number of publications done jointly was collected.

These relationships will be studied using the SNA methodology to calculate the appropriate metrics to measure the network behavior as a whole and of the actors' relationship patterns. These indexes will: (a) identify the importance, influence and interaction degree of each actor or group of actors in the network; (b) map and assess the temporal evolution of the network; (c) analyze the extrapolation of trends of the SNA indicators based on homophily and saturation principles; (d) examine the structural distribution of each network in terms of components, cliques, natural groupings, etc. Still on the perspective of the SNA methodology, will map the attributes of the main actors (division, hierarchical position, scholarship student productivity, CNPq research group, etc.) and performed hypothesis' statistical test to see if and how such attributes can be correlated with their importance in the network.

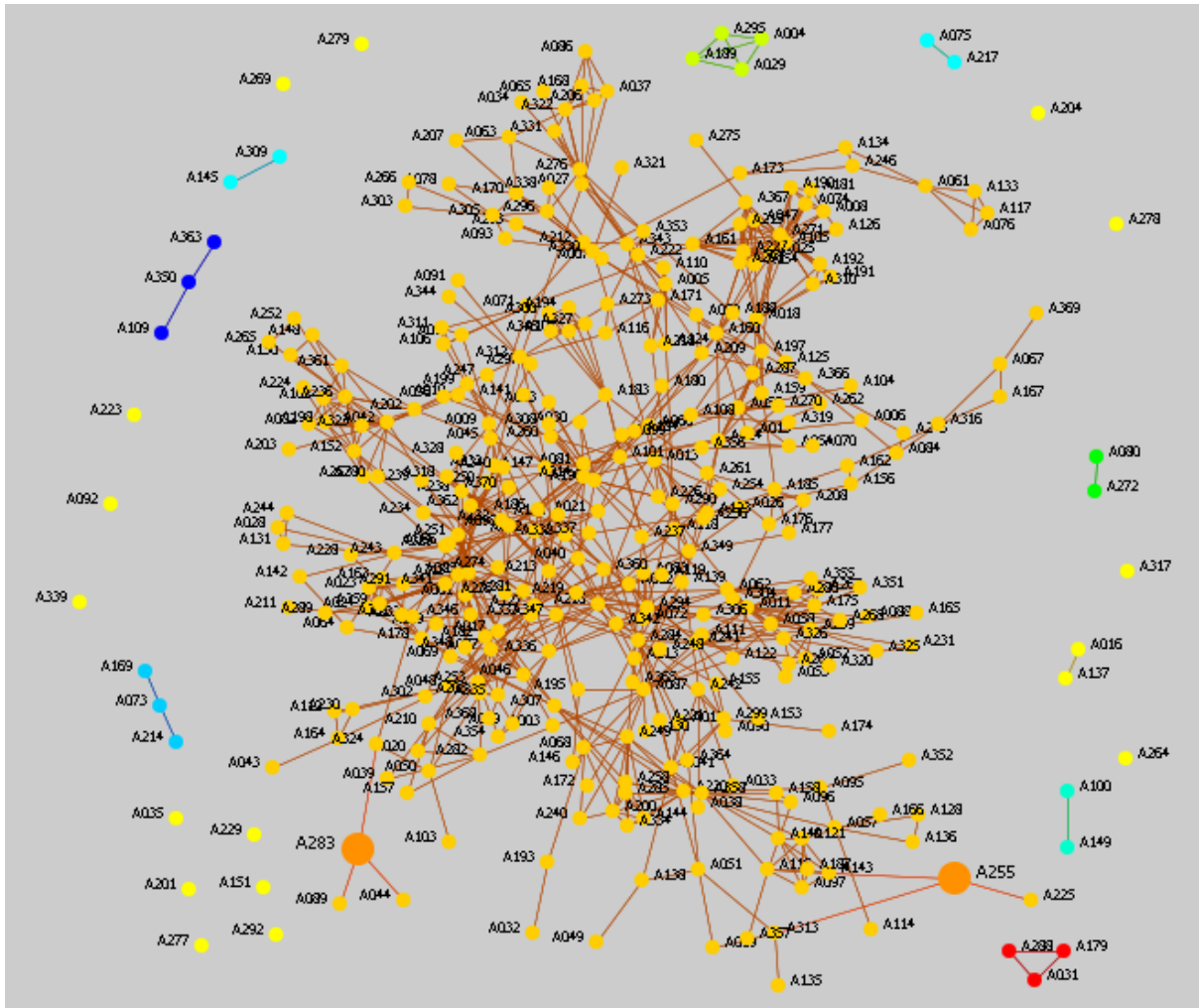
The step that will employ the MEE methodology, data collection will be performed through a survey [11,12] using a questionnaire with closed questions administered by the interviewer to studying the factors that influence the collaboration. This questionnaire is divided into three parts: (a) adapted the model [20], examining behavioral aspects of respondents' relationship, (b) conceptually understand, what characteristics, personal and organizational, influence in the propensity to collaboration with other; (c) obtain indicators to validate the model based on personal choice.

The questionnaire items will be filled by means of Likert-type metric scales [12], ranging from 1 to 5.

#### **5. SOME RESULTS**

The period selected to present some results was the biennium 2007-2008. For this period were selected, through the computer program mentioned earlier, 370 actors (nodes) arranged in a sociomatrix, i.e., an adjacency matrix that specifies the relationships between the actors.

Fig. 5 shows an overview of the IPEN's scientific collaboration network. All network information presented in this article was drawn using the ORA software [4].



**Figure 5. IPEN's scientific collaboration network sociogram - 2007/2008**

Then the network will be analyzed in three complementary ways: first under the aspect of its whole configuration, with the use of global indicators, then under the focus of the key actors, using nodes indicators and finally by observing the groups main of the network.

### 5.1. Global Analysis

A network aggregate overview is presented in Table 1.

**Table 1. IPEN's scientific collaboration network indicators**

Measure	Value
Number of Nodes	370
Number of Links	1746
Density	0.013
Number of Components	25
Number of Isolated Nodes	15
Fragmentation	0.195

The network shows a density of 1.3% over the total number of connections, theoretically possible. At first this may seem a low value, but it has to look more closely, because this indicator assumes that each person could have collaborated and produced some article with all the other 369 people. Now, we all know that it is physically impossible or at least extremely unlikely. What should be done is to adopt a more realistic metric comparing the network density with what would be achieved if everyone had the same collaboration capacity in relation to the individual who collaborated more. Taking the actor with highest total degree shows that he has 23 connections. In this case, the maximum density, physically plausible, would be 0.0623 and, therefore, the network in question has reached 20.9% of this value, which is not a low value.

The network is well connected and therefore a low rate of fragmentation (< 20%). An analysis from the viewpoint of the configuration of components, leading to the same conclusion, to this, notes the distribution of the components size in Table 2.

**Table 2. Components configuration of the network**

Component size (number of nodes)	Frequency of occurrence	Accumulated % of the network
332	1	89.7%
4	1	90.8%
3	3	93.2%
2	5	95.9%
1	15	100.0%

The largest component represents approximately 90% of the network. Therefore, it is a little fragmented network, that is, quite full, since the fragmentation reaches only 10% of the network. In practice, few collaborative networks involving many people can achieve this high level of aggregation.

In this context it is important to note the percentage of nodes that are boundary spanners in the main component. Boundary spanner is an actor who connects his group to other groups. An actor who is positioned as a boundary spanner avoids network fragmentation. Therefore, the removal of this actor will increase the network fragmentation, creating new components.

The network in question, which has low fragmentation and where the main component contains about 90% of the nodes (Table 2), it only makes sense the analysis of boundary spanners for this component. Besides showing the importance of a node, this indicator can also be used collectively to provide insight into the network vulnerability, when computed the percentage of boundary spanners (BSs). However, not all BSs have the same power of fragmentation. In fact, there are some that if removed practically not affect the network, note for example the A283 and A255 nodes in Fig. 5. The most important BSs are usually those who have a high degree centrality in other indicators. For the analysis of the network as a whole, just note that the main component has 53 BSs, i.e., 15% of the nodes, but these only 10 (2.8%) or at most 15 (4.2%) would have real power to fragmentation, as it can be deduced from tables 3 and 4.

**Table 3. The most prominent actors in the IPEN's scientific collaboration network**

<i>Degree centrality</i>		<i>Betweenness centrality</i>		<i>Clique count</i>		<i>Eigenvector centrality</i>	<i>Boundary Spanners</i>
<i>Autor</i>	<i>Value</i>	<i>Autor</i>	<i>Value</i>	<i>Autor</i>	<i>Value</i>	<i>Autor</i>	<i>Autor</i>
<b>A127</b>	0.062	<b>A284</b>	0.191	<b>A098</b>	14	A004	<b>A011</b>
A105	0.049	A022	0.126	A022	11	A016	<b>A027</b>
A271	0.049	<b>A127</b>	0.108	<b>A127</b>	11	A029	<b>A042</b>
A025	0.046	<b>A312</b>	0.105	A251	9	A031	<b>A048</b>
<b>A098</b>	0.046	<b>A068</b>	0.095	<b>A284</b>	9	A073	<b>A056</b>
<b>A284</b>	0.046	A216	0.077	A018	8	A075	<b>A068</b>
A018	0.043	A249	0.076	<b>A236</b>	8	A080	<b>A098</b>
A022	0.043	A160	0.072	A132	7	A100	<b>A127</b>
A251	0.043	<b>A330</b>	0.07	<b>A219</b>	7	A105	<b>A219</b>
<b>A011</b>	0.038	A081	0.067	<b>A314</b>	7	A137	<b>A236</b>
<b>A042</b>	0.038	A013	0.063	A025	6	A145	<b>A259</b>
<b>A219</b>	0.038	A184	0.06	A041	6	A149	<b>A284</b>
<b>A236</b>	0.038	<b>A042</b>	0.06	A160	6	A179	<b>A312</b>
A160	0.035	<b>A098</b>	0.052	A216	6	A189	<b>A314</b>
A216	0.035	A120	0.052	<b>A259</b>	6	A217	<b>A330</b>
<b>A259</b>	0.035	<b>A027</b>	0.05	A362	6	A271	
<b>A314</b>	0.035	A251	0.049	<b>A011</b>	5	A272	
<b>A056</b>	0.032	A358	0.048	<b>A042</b>	5	A288	
A058	0.032	A018	0.048	<b>A048</b>	5	A295	
A249	0.032	A349	0.047	<b>A056</b>	5	A309	

## 5.2. Key Actors Analysis

The connections number of an actor and their relative positioning in the network defines and shapes its power to influence the network performance as a system. To better understand these possibilities, many indicators have been designed. In this work, some of these indicators will be used together to detect the key actors which can be regarded as those of greatest prominence [17,21]. The Table 3 presents a range of the most prominent actors in the IPEN's scientific collaboration network, highlighting those who are BSs.

From Table 3 the most important actors were identified using the following rule of aggregate score (AS): one point, for each time the actor appears in the lists of degree centrality (DC), betweenness centrality (BC) and clique count (CL), and a half point each time the actor appears in the list of eigenvector centrality. These results are shown in Table 4, considering only the actors with  $AS > 1$ .

**Table 4. The most important actors according aggregate score rule ( $AS > 1$ )**

<i>AS</i>	<i>Author</i>	<i>DC</i>	<i>BC</i>	<i>CL</i>	<i>EC</i>	<i>BS</i>
3	A018	0.043	0.048	8		
3	A022	0.043	0.126	11		
3	A042	0.038	0.060	5		*
3	A098	0.046	0.052	14		*
3	A127	0.062	0.108	11		*
3	A160	0.035	0.072	6		
3	A216	0.035	0.077	6		
3	A251	0.043	0.049	9		
3	A284	0.046	0.191	9		*
2	A011	0.038		5		*
2	A025	0.046		6		
2	A056	0.032		5		*
2	A219	0.038		7		*
2	A236	0.038		8		*
2	A249	0.032	0.076			
2	A259	0.035		6		*
2	A314	0.035		7		*
1.5	A105	0.049			1	
1.5	A271	0.049			1	

The betweenness centrality characterizes a particular actor as an intermediate agent in the interactions between two others non-adjacent actors, i.e., they aren't directly connected. Consider the following example: suppose that for each pair of nodes in the network is made a count of geodesic paths that exist between them, and then calculated what percentage of these paths that passes through node X, this indicator is the betweenness centrality of the node X.

This indicator is used to detect people who aren't as connected but can act as intermediate agent for many others. If we were talking about an information network, these people would probably be very well informed because they are in the way of much information.

The eigenvector centrality is an indicator of those actors who are very connected to other actors who are central in the network - the higher your link with other actors who are central in the network, greater the eigenvector centrality. This gives another indication of the importance or power of influence of the actor.

### 5.3. Grouping Analysis

Another approach to analyze a network structure is based on the idea of subgroups or groups within a network. The network division into groups or subgroups may reveal an important aspect of behavioral characteristics or attributes of its members. According to [5] subgroup is a fundamental concept to describe the network structure (the other is the centrality).

The groups or subgroups can be analyzed within a narrower scope to examine, for example, the formation of totally cohesive groups as cliques. A clique is a completely dense subgraph, i.e., nodes belonging to this group have all possible connections between them. The clique is defined as the largest subgraph that satisfies the above definition in the region considered, i.e., cannot add more adjacent nodes to any of the elements without losing density. In network analysis, a clique is a set of nodes, all directly connected to each other, and no other nodes with direct common link to any other node [2]. In a network with more than 300 nodes, cliques containing 3 and 4 nodes are common.

It is interesting to compare the situation of the network with a randomly generated network and has the same density. This can be seen in Table 5 where the real network is in blue and the random networks are in yellow. Note that in the random network the formation of cliques is very infrequent. The clique distribution for our network means that there is a strong tendency for dense grouping of people collaborating. This seems to be an indication that belonging to many and large cliques are good for productivity and in future works this aspect is going to be investigated.

**Table 5. Real network X random network**

<i>Real Network</i>		<i>Random Network</i>		
<i>Number of Cliques</i>	<i>Clique Size</i>	<i>Net</i>	<i>Number of Cliques</i>	<i>Clique Size</i>
100	3	A	11	3
40	4	B	17	3
21	5	C	17	3
9	6	D	13	3
5	7	E	8	3
1	10	F	19	3
		G	15	3

*Random networks generated in 26/11/2011*

## 6. CONCLUSION

This paper has fully described the structure of an ongoing research that is investigating how knowledge is created, transferred and preserved at IPEN.

According to preliminary results it was found a robust and sustainable network of scientific collaboration that over the years has established itself as an effective way for the creation, transfer and preservation of scientific knowledge at IPEN.

The use of SNA methodology allowed to obtain the size and structure of the collaboration network, reveal the most prolific actors and groups beyond those that establish more collaborative ties in R&D. However, further studies are being conducted to cover all the research issues described in this text.

Moreover, the adopted approach presents innovative and challenging aspects, since it defines a structural model to try to explain what personal and organizational factors have a greater influence on the propensity of people to form a partnership to collaborate in R&D.

Finally, it is believed that this work can help senior management of CNEN and IPEN, to adopt policies and initiatives to facilitate the construction of a collaborative environment and, therefore, knowledge sharing and creation.

## REFERENCES

1. A. C. O. Barroso, J. S. B. Reis-Junior, C. A. Monteiro, A. Seary, "Research network involving retired experts as a means to keep alive relevant knowledge – the case of IPEN in Brazil," *Procedia – Social and Behavioral Sciences*, **Vol. 2**, pp. 6475-6482 (2010).
2. M. Kilduff, W. Tsai, *Social Networks and Organizations*, SAGE Publications, London & England (2003).
3. J. P. Birnholtz, "When Do Researchers Collaborate? Toward a Model of Collaboration Propensity," *Journal of the American Society for Information Science and Technology*, **Vol. 58**, p. 2226–2239 (2007).
4. K. M. Carley, J. Reminga, "ORA: Organizer Risk Analyzer Software," Carnegie Mellon University, School of Computer Science, Institute for Software Research International, <http://www.casos.cs.cmu.edu/projects/ora/software.php> (2011).
5. K. Chan, J. Liebowitz, "The synergy of social network analysis and knowledge mapping: a case study," *International Journal Management and Decision Making*, **Vol. 7**, pp.19–35 (2006).
6. J. S. Coleman, *Foundations of Social Theory*, Harvard University Press (paperback edition), Cambridge & USA (1994).
7. \_\_\_\_\_, "Social Capital in the Creation of Human Capital," *The American Journal of Sociology*, **Vol. 94** (supplement), pp. 95-120 (1988).
8. R. Cross, A. Parker, S. P. Borgatti, "A bird's eye view: using social network analysis to improve knowledge creation and sharing," *IBM Institute for Business Value*, <http://www.steveborgatti.com/papers/cross,%20parker%20and%20borgatti%20-%20A%20birds%20eye%20view.pdf> (2000).
9. T. H. Davenport, L. Prusak, *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. Ed. Campus, Rio de Janeiro & Brazil (1998).

10. P. F. Drucker, "They're Not Employees, They're People," *Harvard Business Review*, **Vol. 80**, pp. 70-77 (2002).
11. A. C. Gil, *Métodos e técnicas de pesquisa social*, Ed. Atlas, São Paulo & Brazil (2008).
12. J. F. Hair-Jr, P. Samouel, B. Barry, M. Arthur, *Fundamentos de métodos de pesquisa em administração*. Ed. Bookman, Porto Alegre & Brazil (2005).
13. R. A. Hanneman, M. Riddle, *Introduction to Social Network Methods (textbook)*, <http://faculty.ucr.edu/~hanneman/> (2005).
14. C.-C. Huang, "Knowledge sharing and group cohesiveness on performance: An empirical study of technology R&D teams in Taiwan," *Technovation*, **Vol. 29**, pp. 786-797 (2009).
15. S. Lee, B. Bozeman, "The Impact of Research Collaboration on Scientific Productivity," *Social Studies of Science*, **Vol. 35**, pp. 673-702 (2005).
16. S. Ramanadhan, J. L. Wiecha, K. M. Emmons, S. L. Gortmaker, K. Viswanath, "Extra-team connections for knowledge transfer between staff teams," *Oxford University Press*, **Vol. 24**, pp. 967-976 (2009).
17. J. Scott, *Social Network Analysis: a handbook (2. ed.)*, SAGE Publications, London & England (2000).
18. T. A. Stewart, *Capital intelectual: a nova vantagem competitiva das empresas*, Editora Campus, Rio de Janeiro & Brazil (1998).
19. A. Tiwana, *Knowledge Management Toolkit, The (1. ed.)*, Prentice Hall PTR, Upper Saddle River & USA (1999).
20. P. Totterdell, D. Holman, A. Hukin, Amy. "Social networkers: Measuring and examining individual differences in propensity to connect with others," *Social Networks*, **Vol. 30**, pp. 283-296 (2008).
21. S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York & USA (1994).