PRESERVATION OF KNOWLEDGE THROUGH NETWORKING WITH RETIREES

A. C. O. Barroso[1], J. S. B. Reis-Junior[1], C. A. Monteiro[1], A. Seary[2]

[1] Instituto de Pesquisas Energéticas e Nucleares, IPEN - CNEN/SP
Av. Professor Lineu Prestes 2242
05508-000 São Paulo, SP
barroso@ipen.br, jose.sergio.junior@googlemail.com, monteiro@ipen.br

[2] Simon Fraser University
School of Comunication
Burnaby, BC, Canada
seary@sfu.ca

ABSTRACT

Loss of emphasis or phasing down of nuclear programs has reduced substantially the hiring of new employees as a result nuclear organizations, in most countries, are experiencing reduction in the workforce and the average age of their technical professionals are around mid fifties. Knowledge management activities with emphasis on knowledge preservation have become a crucial issue for such organizations.

This work studied a spontaneous knowledge preservation mechanism at IPEN that could be leveraged and may be replicated in other organizations. Crossing examining publications and human resources data base, with some alias detecting algorithm a large collaboration network involving retirees and current workers of IPEN was unveiled. Using simple indicators and advance techniques of social network analysis the following studies were performed: (a) assessment of the network performance; (b) characterization of its key global properties and detailed structure; (c) characterization and assessment of the role of its key actors; (d) analysis of groups and subgroups patterns; and (e) longitudinal (time evolution) of the network and assessment of its robustness. Rich insights came from this study concerning the value of this mechanism for IPEN and also about the essence of the common interest that constitutes the "glue" for such mechanism. While more detailed network analysis will still go on for a couple of months, a new phase has already been started with a formulated conceptual model, consisting of four latent variables and thirty six observable ones, to "explain" at the actor level, in this particular setting what matters when engaging in collaboration. Upon finishing this new phase network data and actors' survey data will be cross correlate to provide a more fully understanding of this amazing mechanism.

1. INTRODUCTION

Nuclear organizations, in most countries, have for many years endured a process of a reducing workforce and of a shift of the average age of their technical professionals towards the mid fifties range. Loss of emphasis or phasing down of their nuclear programs has reduced substantially the hiring of new employees. As a result some knowledge management activities with emphasis on knowledge preservation have been undertaken in many such organizations.

This work is part of an ongoing research effort aimed at studying spontaneously emerged knowledge preservation mechanisms at IPEN that could be leveraged and may be replicated in other organizations. In a previous paper recently presented at the XXIX Sunbelt Social

Networks Conference a first round of results were presented, with emphasis on: (a) measuring the output of the network; (b) characterization of its key global properties and those of its key players; and (d) test of its robustness. This work shows some further results along the same path and starts a new perspective trying to unveil the factors and mechanisms that were responsible for the creation and nurturing of this network.

2. THE CASE STUDY

At IPEN, the largest research institute of CNEN - Brazilian Nuclear Energy Commission - a social network, SN, involving retired experts has emerged naturally and is helping to transfer and keep alive the knowledge that otherwise would be lost by the institution. Enabling conditions that presumably made possible the formation and sustainability of this network can be summarized as:

(a) there is a career progression rule stating that upon completion of master and doctor degrees IPEN's employees are automatically awarded a significant salary increase;
(b) IPEN is sited at São Paulo University – USP – campus and IPEN's researchers, with doctoral degrees and upon meeting some criteria, can teach graduate disciplines within the USP grid;
(c) as a result of the above, all technical employees are naturally encouraged to pursue graduate studies within IPEN/USP grid of disciplines and research themes;
(d) the Brazilian Council for Scientific and Technological Development - CNPq - offers a fellowship that includes some monthly income for researchers who achieve a certain sustained level of scientific productivity;
(e) there are a good number of "medical and psychological" reasons for a person to stay intellectually active after retirement;
(f) office space and access to laboratories and other institute facilities are provided by IPEN to retirees upon a special "voluntary work contract arrangement";
(g) IPEN's site, inside São Paulo University campus, is kind of an oasis in a busy and overcrowd city like São Paulo; and
(h) the presence of CIETEC, the largest technology based business incubator of Brazil, inside IPEN's site.

The net result is that are many intangible and a few tangible incentives for productive retired IPEN researchers to voluntarily keep their professor status and their respective research activities. Also with a high probability they will be advisors of younger IPEN employees feeding the network that we have analyzed. Tangible benefits materialize for researchers previously awarded with CNPq productivity fellowship as this network makes possible for them to continue publishing to maintain the fellowship. Other eventual and less probable benefits are partnership or eventual consultancy for entrepreneurs' firms sited at CIETEC.

Social network analysis (SNA) is a process for mapping and studying the network of relationships among people, teams, or across organizations and communities. SNA is particularly effective in assessing the flow of information and knowledge through communication and collaboration. It also identifies people who are central and peripheral to such process.

As a field of knowledge, the methodological framework of SNA has emerged as a means to fit the needs of special studies in anthropology, sociology, social psychology. Most of the

ideas were developed from the thirties to the seventies, but it was from the late seventies to the nineties when mathematical techniques and tools became available and software was developed to make easier the endeavor of the analysts. As a technique, SNA became increasingly popular in the last ten years when practitioners and scholars from fields as diverse as epidemiology, physics, economics, business and administration and knowledge management became especially interested in its applications. A nice summary of the origins and development of SNA can be found in [1].

Most of the mathematics used in SNA comes from graph theory and matrix algebra. To those interested in the subject, the texts of [2] and [3] are excellent starters; the former is easier for fast reading and the latter is denser in the theory and mathematics of the methodology. For those not familiar with the field, we will use a few lines to introduce a minimum of the initial concepts and later, during the analysis, comments will be added as necessary, in such a way that the reader can get an understanding of the work here reported. The use of mathematical expressions will be avoided in this paper, despite some loss of precision, because the intended readers are not SNA experts but nuclear engineers and technology managers. Instead the concepts and ideas behind the analysis will be presented in simple (colloquial) terms.

Given a population of actors and one relation of interest to be studied, the set of actors and all the (directional) ties showing the existence (and the intensity) of the relationships among them constitutes a network. So if the actors are labeled by numbers, e.g. i and j, the existing relationships can be represented by ordered pairs (i, j) with i and j varying from 1 to N, where N is the total number of mapped actors (population). Such representation has an isomorphism with a directed (valued) graph and also with a matrix. The matrix is usually known as socio-matrix or adjacency matrix and the graph is sometimes called a sociogram, as a leading pioneer of SNA Jacob Moreno has named this network metaphor. Fig. 1 below illustrates this concept for a very simple network of 4 actors.
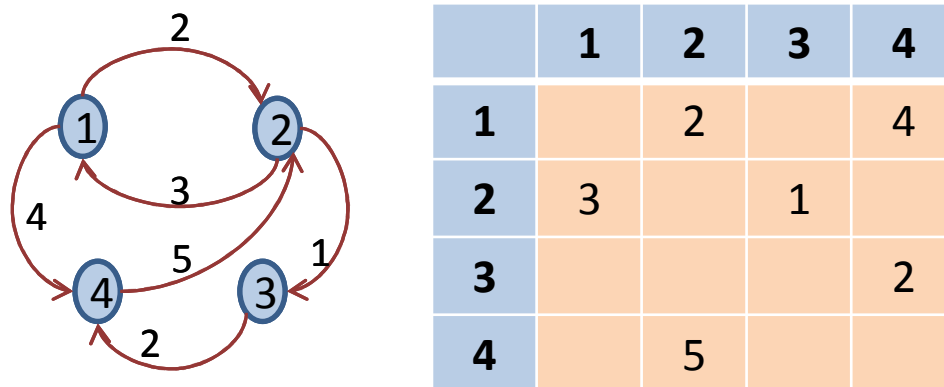

	1	2	3	4
1		2		4
2	3		1	
3				2
4		5		

Figure 1. A simple valued directed graph and its corresponding adjacency matrix.

The studied population consisted of retired IPEN workers and any IPEN current worker whose collaboration with the former has resulted in a publication of any kind. Based on

IPEN´s database of publications, co-authorship was used to define the relation of the network definition and data was collected annually from 2001 up to 2008. The social networks resulting from co-authorship were grouped into two-year periods and analyzed.

From the IPEN publication database, xml files containing publication characteristics, authors name and affiliation, were generated for years 2001 till 2008. From the Human Resources Database, corresponding csv files for active and retired employees were also obtained. An algorithm was developed and a script was coded with a two-fold purpose: (a) to identify alias citation in the publication files and reduce all of them to the standard format of the HR Database generated files; and (b) to generate input files for Excel, and SNA softwares (Agna, NetDraw, UCINet, Pajek and MultiNet). The adjacency matrix is symmetric since it was assumed that collaboration in a paper is inherently a bidirectional relationship. The strength of the links between peers was measured by the number of publications in which they have collaborated. In other words its graphical representation yields an undirected valued graph.

3. ASSESSING THE NETWORKS' OUTPUT

A set of four aggregated indicators were defined for this purpose. Their definitions are presented below and their values and other pertinent data are summarized in Table 1.

I_1 – participation index of the average retiree, calculated as the number of publications involving retirees / total no. of publications);

I_2 – partnership index, measuring the "social capillarity" of the average retiree, computed as the ratio of the number of non-retired co-authors to the total number of retired co-authors;

I_3 – productivity index of the retired research, described by the ratio of the number of publications with participation of retirees over the total number of them;

I_4 – productivity index of IPEN researchers (retired or not), given by the total number of publications over the total number of authors.

Table 1. Overall data and indicators of the retirees' participation

Period	Publications of IPEN		Authors of IPEN			Indicators			
	Total	With retirees	Total	Co-authors of retirees	Retirees	I_1	I_2	I_3	I_4
2001_02	1486	268	390	130	41	18.0%	3.17	6.54	3.81
2003_05	1200	178	319	91	34	14.8%	2.68	5.24	3.76
2005_06	1254	179	349	104	32	14.3%	3.25	5.59	3.59
2007_08	1530	186	361	121	34	12.2%	3.56	5.47	4.24

As one can see, retirees' contribution is significant and publications involving them range from 12,2% to 18,0% of the total. Each retiree collaborates on the average with 3.02 non-retired researchers. Finally, comparing indicators I_3 and I_4, one observes that the average productivity index of the retirees is significantly (29% to 72%) better than that of the joint community of researchers. It also must be noted that the participation and the capillarity

indexes are really very good. In a previous publication [4], time slices of one year were considered to observe the network and, at that time, when discussing the capillarity (partnership) index it was conjectured that such period could be inappropriate since on the average it takes some two years of research work before publications emerge from it. The values for this index are 50% greater than those previously obtained, a fact that corroborates the previous argument.

4. CHARACTERIZING THE NETWORK

For most of the analysis Ucinet 6 [5] has been used, however for some simple calculations Agna 2.1 [6] was chosen. The former is much more powerful and resourceful, while the latter is easier to use.

4.1. General features

The first thing to look is the connectedness of the network configuration in terms of components. The ideal situation would be to have just one whole network, which means that everybody is reachable in a walk through the network; however in a large collaboration network some degree of fragmentation is usual. It is common to have many components that do not communicate with each other, which means people that collaborate with each other but only with in their own group (component). From a whole integral network to a very atomized one there quite a variety of possibilities and, to get a feeling of the configuration connectedness it is possible to define some figures of merit by counting the number of independent components that are needed to account to 70%, 80% and 90% of the population. The smaller these numbers the more connected the configuration is.

Fig. 2 presents, for the four periods analyzed, the graph of the cumulative percentage of the population accounted by successively adding the components from the largest ones to the smaller ones.

As one can see, in most cases 70%, 80% and 90% of the population are achieved with respectively 2, 4 and 8 components.

Some general indicators are presented in Table 2. Diameter is the length of the longest geodesic in the network. Geodesic is simply the minimum number of links one has to walk to go from actor to another, so diameter is the length of the geodesic that connects the actors that are farthest apart.

Density is a measure of how connected the network is and it is a ratio of the number of existing links to the maximum number theoretically possible. Since an actor can in principle be linked with N-1 actors, the maximum number of links grows with N^2. However depending on the type of the relation a person has a physical limit to establish relationships, and as the population grows networks tend to become less dense.
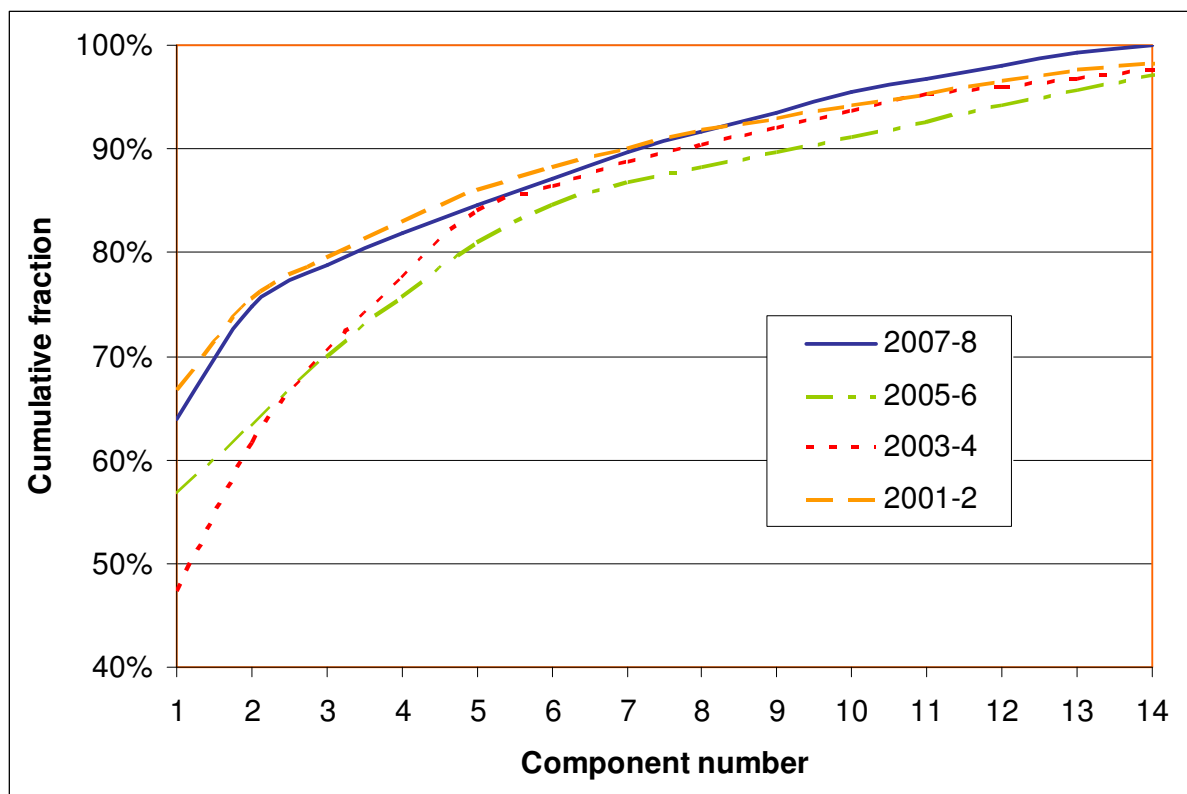
Figure 2. Component size distrubution.

For this reason we have introduced a naive casuistic indicator of the maximum feasible density (for a given population and relation type) and this indicator is calculated as the density of the network if all actors were as connected as the leading one. Then the density ratio of the actual density to the maximum feasible one can be used a measure of how well the network is fulfilling its connectedness potential. It is still a demanding indicator if one notes from Table 3 that for the period 2001-02 the most connected actor has collaborated with 20 different persons, so the number for this period says that the network is exploiting 21% of its maximum feasible connectedness capacity. It is good to see that this indicator is evolving upwards as it has climbed from 21% to 26.3%, however it is not expected that it could go over some 30% because of statistical reasons. Since the ability to collaborate productively with many persons depends on personal traits and those are not evenly distributed across a population, by choosing the maximum the indicator is comparing all the population with someone that is at least some 3 standard deviations (σ) over the mean with respect to those desirable traits.

It has to be remarked that density and diameters are calculated as if the network were completely integral (just one component) and as it is shown in Fig. 1 this is not the case. Therefore, the way the software performs the calculations, the result will yield the diameter of the largest component, which is a benevolent estimate. On the other hand, density calculation includes all nodes even if they do not belong to the same component, which is certainly smaller than the density of the largest component. An analysis of the 2 largest

components of each period, as a table with the component size distribution was performed but not included here.

Table 2. Some global indicators of the network photos

	2001_02	2003_04	2005_06	2007_08
Diameter	9	7	9	11
Density	0.0248	0.0305	0.0334	0.0307
Max feasible density	0.1176	0.1129	0.1343	0.1169
Density ratio	0.2105	0.2701	0.2490	0.2631

Looking into the network, it is interesting to check on the general profile of collaboration and data from Table 3, although not self-explanatory, give some insight in this respect. Considering the period 2007-08, for example, one can note: 1 (one) outsider, 2 (two) most collaborative individuals that have collaborated with 18 persons, 34 (thirty four) individuals that have collaborated with 4 others making up the distribution mode, and a mean of 4.7 collaborators per element. The proximity of mean and median indicates a typical well behaved distribution whose peak (mode) is smaller than both the mean and the median.

Table 3. Overall collaboration profile

Distribution of the Number of Collaborators per Element (self excluded)				
Period	**2001-02**	**2003-04**	**2005-06**	**2007-08**
Minimum	(0, 4)	(0, 6)	(0, 4)	(0, 1)
Maximum	(20, 1)	(14, 1)	(18, 2)	(18, 2)
Mode	(1, 30)	(2, 28)	(1, 24)	(4, 34)
Mean	4.21	3.79	4.51	4.74
Median	4	3	4	4
Authors / publication*	2.56	2.31	2.60	2.58
(x, y): x = no. of collaborators; y = frequency of occurrence. * additional information				

More noticeable is the fact that mean number of partners per element shows an increase of about 12% across the four periods, which is significant and shows that the network is becoming denser. This growth is really due to the increase of new productive links and not just because more people are collaborating on the same number of papers per person, because as one can see the average number of authors per publication varies less than 2%. In other words the ratio of the elements on lines five (mean) and seven (authors/publication) is kept nearly constant during all the periods.

Another feature that was assessed was how long a productive retiree stays in this network, which is a contributing factor for the sustainability of the network. This distribution of postretirement engagement and production will be discussed in another publication, but, just to exemplify, it is worth noting that a very linked retiree has been in the network for over 22 years.

Taking into everything that was discussed so far there are no doubts about the significance of the benefits for IPEN of this retirees centered network.

4.2. Looking for Patterns

4.2.1. Leading actors at each period – degree centrality

A very basic indicator of centrality of actor is called degree, which for these networks means the number of links a given node has. Ultimately degree is measure of social performance with respect to the relationship under analysis and since we dealing with valued graph, it is also worthwhile to use the weighted degree where the value of the link is included in the sum. This means that if actor **i** has 3 publications with actor **j** their link will be counted as 3.

Both degrees and weighted degrees were computed for all periods and a list putting together the 7% top weighted degrees of each period amounted to 30 actors. It interesting to note that: (a) all the 20 that made up the core list (see the section on leading groups) were included; (b) eleven (11) retirees were included in this list that is 37%; (c) 80% of actors of this list (top 7% weighted degrees) were also present in a similar list for the top degrees. Tables were not included because they would consume a lot of space.

Many other centrality indicators were computed, but their joint analysis are beyond the purpose of this article.

4.2.2. Leading groups

4.2.2.1. Cliques

In a network graph, a clique is a maximum complete subgraph with three or more nodes. This means a set of nodes (actors) in which every one is adjacent (tied) to everyone else and there is not any other node that satisfy this requirement of adjacency to all members. Cliques indicate very cohesive subgroups that somehow leverage the influence of its participants. In a larger network one can find many cliques of three and four elements and if larger cliques are not found this can be a sign of atomized small groups, which could hamper the capacity of more extensive and orchestrated collaboration through the network. A clique analysis was performed focusing on large cliques (≥ 6 elements), the results are summarized on Table 4 and briefly commented after it.

Table 4. Large cliques

Period	Size of the network	Cliques of					Total
		6 nodes	7 nodes	8 nodes	9 nodes	10 nodes	
2001_02	171	6	1		1		8
2003_04	125	1	2				3
2005_06	136	4	5		1		10
2007_08	155	5	2			1	8

A quite satisfactory distribution is presented at every period, meaning that a large portion of the population is participating of these very cohesive subgroups. Also the configuration seems to be improving with time, as these subgroups seem to "open" to new members. A deeper analysis of how these cliques are evolving with time is presently being undertaken.

4.2.2.2. Core and periphery

The structural notion of core and periphery is common to social network analysts as well as to many other researchers and analysts of other fields. In politics very often people talk about the hard core and the lower clergy of a political party or of a coalition of parties. This concept is a metaphor for an idealized network configuration with a core group in which every one has links to each other and a peripheral group of nodes that have no direct links among themselves and can only communicate through the core. This would yield a socio-matrix with four distinct regions: (a) a fully dense (except for the null diagonal) one representing the core, (b) a fully empty one representing the periphery and (c,d) two sparse regions with the connections from core to periphery and vice-versa. Of course real networks are almost never like that that but one can try to different kinds of permutation of the socio matrix to get as close as possible to this pattern. As this is done, core-like and periphery-like groups are recognized and also some parameters to quantify the goodness of the fit of this structure to the real network. It was [7] who have in a more systematic and sound way produced the techniques to perform such analysis.

Such analysis was performed using both the categorical and continuous partition and the results summarized on Tables 5 and 6. It must be noted that the goodness of the fitting was not so good, which means that our network is somewhat far from a core-periphery structure. In fact, using the categorical method, for the fit to be reasonable in one case the group would be much greater than the periphery. The meaning is that this collaboration network has very democratic practices of access to the more active groups.

Table 5. Core/periphery partition

Number of elements in core/periphery to the data network								
Period	2001-2		2003-4		2005-6		2007-8	
	Core	Periphery	Core	Periphery	Core	Periphery	Core	Periphery
Categorical	5	166	20	105	119	17	5	150
Continuous	5	166	6	105	7	129	3	152
Adopted	5	166	6	105	7	129	5	150
Retirees in core	1		2		1		3	

Table 6. Core participation along the periods ("dance of the chairs")

Core elements	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
2001-2	x									x		x							x	x
2003-4		x			x		x		x					x		x				
2005-6		x		x	x			x							x		x	x		
2007-8							x		x		x		x	x						
retirees highlighted in yellow																				

It is interesting to observe that no one has been present in the core for more than two periods and in fact the core members of 2001-02 have not come back to the core again, which confirms the democratic structure of more or less active participation in the network. One might argue that the size reduction of 2007-08 is indicative of a concentration, but the effectiveness of the fit was not so good, meaning that the peripheral members were more linked among themselves than in a typical concentrated structure.

Table 7. Retirees presence: core vs. whole network

Period	Network actors			Core members		
	Total	Retirees	% of Retirees	Total	Retirees	% of Retirees
2001-2	171	41	24.0%	5	1	20.0%
2003-4	125	34	27.2%	6	2	33.3%
2005-6	136	32	23.5%	7	1	14.3%
2007-8	155	34	21.9%	5	3	60.0%

Finally Table 7 put into perspective the retirees' presence in the core. As one can see, except for the last period the presence of retirees is entirely compatible with their fraction of the population. In fact one could expect a larger presence, since the definition of the network was retirees and people who have collaborated with them. This is indicative of the fact that: (a) retirees collaborate much more with non retirees than with themselves; and (b) some of the people collaborating with retirees are also very active researchers with a lot of links to the more peripheral persons. Both facts can be confirmed by the leading actors analysis presented

in a previous section and by a detailing study of personal profile of these actors. All these findings are very welcome because they show that transfer of knowledge is taking place and that a change over of roles of leading researchers in the network is happening.

4.3. Peeking Inside the Network

For small networks, just by looking into the graph it is easy to extract most of the "hidden" information, however with more than 40 nodes things get too complicated to be seen and indicators and mathematical techniques can give more insight. Graphical exploration can be very much improved with the help of mathematical algorithms and statistical analysis. A handful of these resources can be found in MultiNet [8] and it can be very helpful for detailed analysis of the network structure. Fig. 3 illustrates what we are doing in this respect.
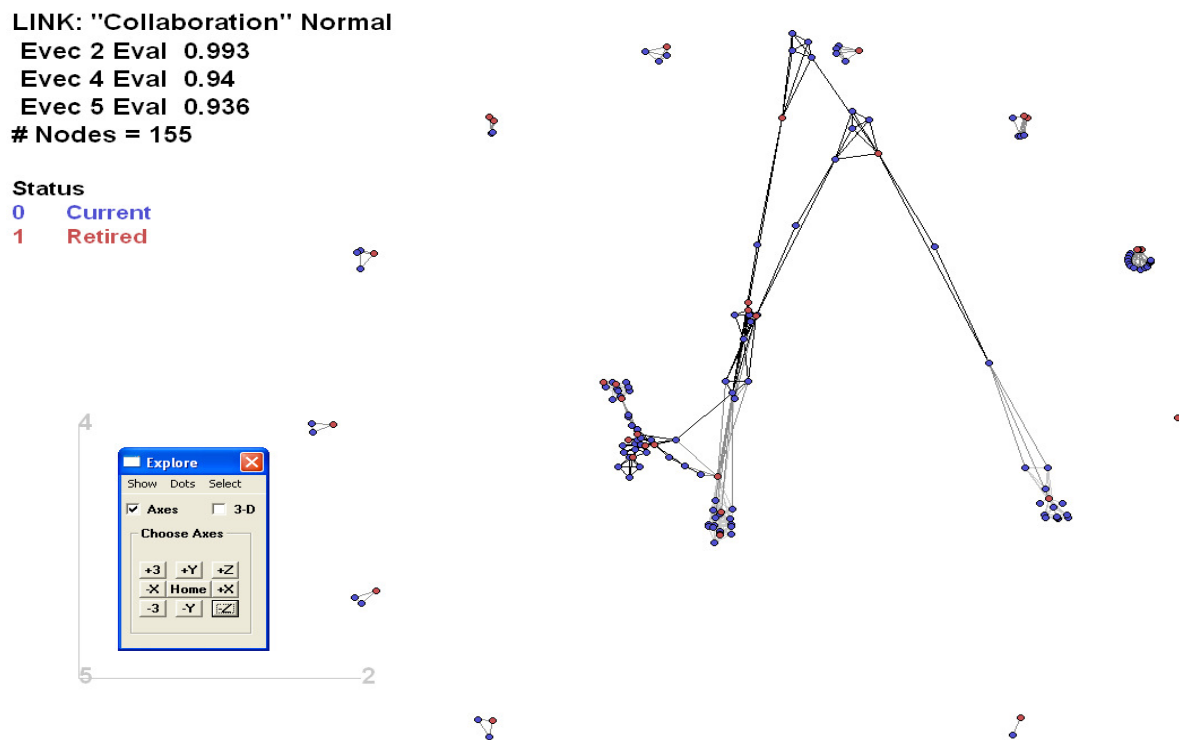

Figure 3.   The 2007-8 collaboration network explored with multinet.

The ortho-normalized eigenvectors of the adjacency matrix were used to get a 3-d picture of the network; distances between the nodes were made proportional to their coordinates in this referential system. We have chosen the 3 vectors that yield a better image for the purposes we had in mind at the moment of this analysis and then we magnified the distances only for the main component. With this type of picture one can spot a lot of interesting information, as for instances actors that, although not so connected (active), function as bridges between densely connected subgroups. After selecting the function information, one just has to point at any dot

(node) and the software gives us all the node attributes and descriptions of the links to and from the selected actor.

5.   INTERVIEWING KEY ACTORS

We have formulated a conceptual model to explain, in this particular setting, what influences people in their collaboration choices. After browsing through the literature and brainstorming with some colleagues a series of variables were identified. A couple of interviews with key retirees were conducted to get some confirmation of the initial list. Then they were refined and rearranged in a construct made up of four factors or latent variables and their respective observable (or proxy) variables. To further refine and validate this construct, a series of semi-structured interview are now being conducted with the most frequent collaborators of the key retired actors. Upon finishing this step a questionnaire will be applied to the whole network.

As of now our model has four latent variables which are: (a) cognitive factors – characteristics that matter if the choice of a collaborator were to be done in pure rational basis; (b) emotional factors – personal traits that make enjoyable joint work; (c) cultural factors - issues related to values and basic assumptions that influence behavior; and (d) circumstantial factors – existing conditions of a given time that can facilitate or catalyze the collaboration. These latent variables have, respectively, seven, seven, five and nine observable variables under them. Upon analyzing the fifteen interviews that are being performed we expect to reduce this menu of variable to simplify the questionnaire.

6. CONCLUSIONS

So far the potential and robustness of the studied network has been satisfactorily demonstrated. It is a long going and sustainable network whose main driver mechanisms are already emplaced and working properly at IPEN.

The question if this mechanism can be replicated, partially or in totality, in a different setting is still open. The current research expects to provide some answers in this respect. The variables that are important will be assessed and correlated with the network performance. We hope to develop similar studies in other research institutes of CNEN. In a future step a "replication model" can be developed and tested in organizations that can provide the appropriate boundary conditions. From point of view of IPEN, there is no doubt that this is an invaluable mechanism to preserve relevant knowledge.

REFERENCES

1. J. Scott., *Social Network Analysis – A Handbook (ed. 2)*, Sage Publications, London, England (2000).
2. R. Hanneman, M. Riddle., *Introduction to Social Network Analysis*, *www.faculty.ucr.edu/~hanneman/nettext/* (2005).
3. S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York, USA (1995).
4. A.C.O. Barroso, J.S.B. Reis-Junior, K. Imakuma, M.O. Menezes, A.J.Seary, "Research network involving retired experts as a means to keep alive relevant knowledge – the case

of IPEN in Brazil," *Proceeding of Meeting*, XXIX Sunbelt International Social Network Conference, San Diego, CA, USA, March 10-15, (2009).

5. S.P. Borgatti, M.G. Everett, L.C.Freeman, Ucinet for Windows: Software for Social Network Analysis, Harvard, MA: Analytic Technologies (2002).

6. M.I. Benta, "Studying Communication Networks With Agna 2.1," Cognition, Brain, Behavior, **Vol. IX (3)**, pp.567-574 (2005).

7. S.P. Borgatti, M.G. Everett, Models of core/periphery structures, Social Networks 21 (1999) 375-395, http://www.analytictech.com/borgatti/publications.htm.

8. A.J. Seary, W.D. Richards, MultiNet for Windows 5.01. Program for interactively analysing and visualizing large sparse networks with integrated Negopy Fall 2007, http://www.sfu.ca/personal/archives/richards/Multinet/Pages/multinet.htm.