Social Networking, 2014, 3, 134-141
Published Online February 2014 (http://www.scirp.org/journal/sn)
http://dx.doi.org/10.4236/sn.2014.32017

Studying Group Dynamics through Social Networks Analysis in a Medical Community

Ruben P. Albuquerque[1], Jonice Oliveira[1], Fabrício F. Faria[1], Rafael Monclar[2], Jano M. de Souza[2]
[1]Graduate School in Computing Science (PPGI), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil
[2]Systems and Computing Engineering Graduate School (COPPE), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil
Email: jonice@dcc.ufrj.br, rrpero@ppgi.ufrj.br, firminodefaria@ppgi.ufrj.br, rastumon@cos.ufrj.br, jano@cos.ufrj.br

Received December 26, 2013; revised 28 January 2014; accepted 19 February 2014

ABSTRACT

In 2008, the Brazilian Department of Science and Technology created the INCTs (Brazilian Science and Technology Institutes). One of them was the Cancer Control INCT. Due to its importance and considering that there are different groups working together in the same area, it is important that they collaborate intensely. Envisioning an empowerment of scientific collaboration, the BRINCA project was created to support a set of analyses of the social networks from this particular INCT. These analyses were created by mining curricular and publications bases, and identifying different types of scientific relationships and areas. We were able to observe, for instance, how the interaction is amongst researchers from related areas, which researchers were more collaborative and which ones were isolated from the network. These analyzes were used by the INCT coordination to understand and act to improve scientific collaboration.

KEYWORDS

Social Networks; Scientific Collaborations; Data Mining

1. Introduction

The Brazilian Government created the National Institute of Science and Technology (INCT) to minimize the division and disintegration that exists amongst scientific groups. The proposal is to join different researchers, universities and research groups of excellence, in Brazil and abroad. One of these institutes is the Brazilian Institute of Science and Technology for Cancer Control [1] that is controlled by the National Institute of Cancer (INCA).

In this scenario, the BRINCA project (Balancing and Analyses of Scientific Social Networks in Cancer Control) was created. The main goals of this project are to analyze how the Cancer Control INCT members collaborate and how the scientific knowledge flows amongst the different researchers and institutes, and the members of the group.

An important aspect of our project is the temporal analyses, understanding the network evolution over the years, including important research areas and when they became more relevant. To enable these analyses, a computational environment was built to support the collection and interpretation of historical data, as well as the identification of possible problems in group dynamics.

This article consolidates and extends the seminal results [2] that were presented at the first Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), a satellite event of the XXXII Brazilian Computer Society Conference in July of 2012. In this article we briefly describe the recent works in the field of medical social networks (Section 2). In Section 3 we detail our proposal, the BRINCA project and its current results in Section 4. In addition, we present related works (Section 5) and conclude this work, pointing to some future work paths (Section 6).

2. Social Network Analysis in Medicine

Social network analysis in medical context is two-fold. First, it is used to contain disease dissemination, to pre-

vent it from achieving an endemic or epidemic level. This can be made through analyzes in the social networks of those infected and predicting how the disease can spread [3-5]. The second usage is the identification of expert networks [6,7], which is the focus of this work.

3. BRINCA Project

The BRINCA Project aims to map the knowledge exchanged amongst Cancer Control INCT researchers, as well as identify how groups develop their research efforts and how professionals interact with each other. So, this project aims at the identification of scientific social networks, the provision of mechanisms for complex analyses to obtain an improvement in the collaboration amongst the main specialists.

The reports provided can help to detect weak or strong points in the interaction between research groups, centres, and countries, assisting in the guidance of scientific development and funding politics [8,9].

In next topics, we describe details of our approach for the analysis of the INCTCC social network.

3.1. Architecture

The architecture developed for our work has its steps shown in **Figure 1**.

The data sources are Lattes [10] and PubMed [11], which will be presented with more details in Section 3.2.

We use Kettle [12] to orchestrate the extraction, treatment and cleaning routines. The visualization layer is composed by Gephi [13], Tableau [14] and independent reports. The metrics were calculated by Gephi [13] and stored in aData Warehouse. Section 4 details the visualizations and analyses.

3.2. Data Sources

The Lattes Curriculum is a Brazilian nation-wide curricular database with all the curricula of scientific professionals in Brazil. All of these curricula were downloaded by XML-Lattes Tool [15] and PubMed data from its own Web service interface.

After the data extraction, transformation and loading processes, our data warehouse (multidimensional database) stores different types of relationships between two researchers over time. The scientific types of relationships are:

- Project Participation—being member of a project team;
- Co-authored—two people work together in a publication;
- Advisory work—a professor supervises a student's work;
- Examination board participation—professors who participate in a committee, to judge and evaluate a thesis;
- Judgment commissions—professors who participate in a committee, to judge and evaluate scientific work—as publications (programme committee), project proposals—or evaluate candidates in hiring processes; and
- Other types of scientific production (e.g., patents).

In addition to relationships, each one of the researchers has an individual profile, built with one's personal attributes, such as: Academic Level (PhD, MSc., or BSc.); Research and activity area; Number of Publications (per type, such as journals, proceedings, technical reports, …); Number of Project participations; Number of Thesis Advice participations; and Number of Participations in Examination Boards. Research and activity areas indicate what areas a researcher is connected with. Examples of research and activity areas are HPV and thyroid cancer.

3.3. Multidimensional Model

All the details of scientific interactions, such as type, frequency, and members of a social network (and their profile) are stored in a Data Warehouse, which obeys a multidimensional model, shown in **Figure 2**.

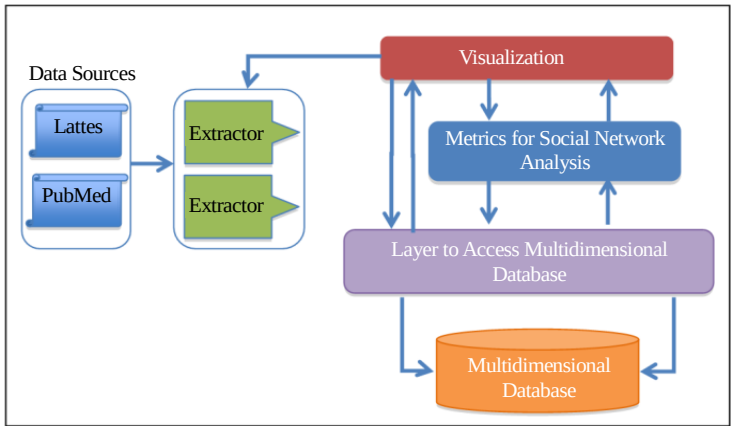

Figure 1. BRINCA's architecture.

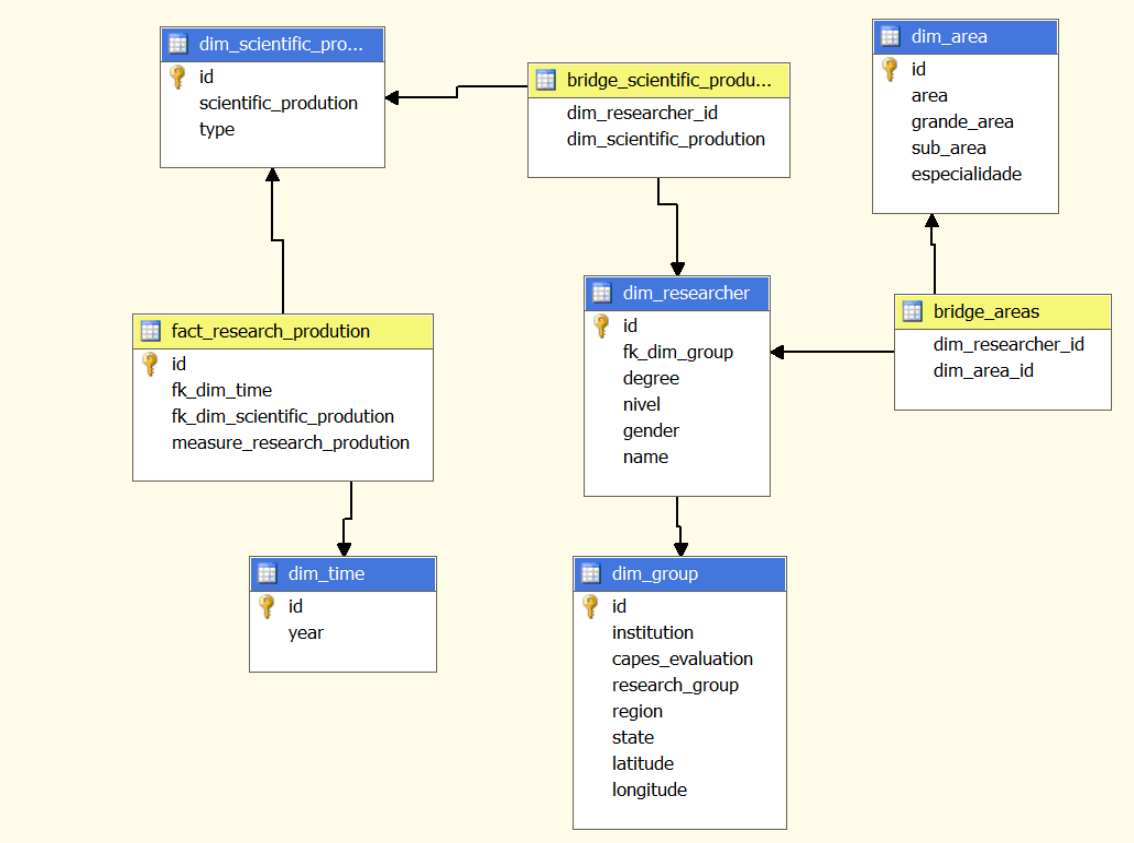

Figure 2. Multidimensional model.

This model has a fact table that aggregates the scientific production per year, via an association with the Time and Scientific Production dimensions. The Scientific Production dimension represents each production made by one or more researchers (Researcher Dimension), who can participate in groups (Group Dimension). The Group Dimension is related to Research Groups and has information on its evaluation and location. All researchers can have one or more expertise areas (*i.e.* Genetics, Biochemistry, etc.).

Based on this model and using our analysis tools, we were able to get the results presented in the next section.

4. Current Results

The main issue of this project (described in Section 3) is to understand the interactions amongst researchers, and the role of Cancer Control INCT in the promotion of scientific cooperation in Cancer.

The works developed in the Cancer Control INCT are classified as per research themes. For each theme, there are sub-projects [1], which has researchers associated to them. Project members can be researchers with the INCT, and also from other (domestic or foreign) institutions. To provide the results below we used data from 122 researchers, without introducing the students involved in the subprojects.

One of the analyses points the most connected researchers in the network. A researcher with a high degree of relationships can be a person with a high level of influence or specific expertise, not always with a supervising position as department managers or project leaders. The relationship average network is 8.496. In a big network, it is usual to have subnets. The relationship average of the most connected nodes in a subnet is 2.667. Some nodes, with a higher linkage degree, are shown in **Figure 3**. Red nodes are department or project heads.

From the 122 researchers, 8 of them are people with no connection with other INCTCC researchers, although they have external links. That is, they are nodes disconnected from the whole network, as seen in **Figure 4**, which shows members and main research area (colour). However, in **Figure 4** two researchers are not counted as they are not associated with any area, showing only 6 disconnected nodes in Medicine (green node), Veterinary Medicine (red), Pharmacology (pink), Pharmacy (blue) and Computing Science (purple).

The network has 8 researchers who act as “bridges”, connecting large groups. Amongst these 8 researchers, 6 are central nodes (with no higher connection degree). The metric used was betweenness centrality. The detection

of bridges is important for us to verify the weak points of our network, which can be rendered fragile and could be easily divided into subgroups if a member, who is a bridge, leaves the group.

Figure 5 shows a piece of the INCTCC's social network, where 14 clusters were identified through Modularity metric, but only 3 could be associated with INCTCC areas. They are: Medicine (20.51% of the researchers), Collective Health (16.24%), and Genetics (9.4%).

Analyzing internal and external interactions, we identified 12 researchers (9.84%) with a greater number of connections with external researchers (who are not members of the INCTCC), compared with only 3 (2.46%) that have strong internal connections with INCTCC researchers. Having a "very intense connection" or "very strong connection" means a node, whose frequency of interaction with other member exceeds 70% of the highest interaction frequency in entire network, which is of 30. Any node with more than 21 interactions with another can be considered as having a very strong relationship with him/her. We identified 6 researchers (4.91%) with very strong relationships with other researchers, who act in different areas.

Since the creation of the INCTCC in January 2009, with its official implementation in June that year, the social network changed. **Figure 6** shows co-author relationships in INCTCC pre-creation and during its development.

These changes affected the average degree of the network. Analyzing only the co-authorship relation, we have the following as average degrees: 1.009 (2007), 0.957 (2008) 0.410 (2009) 0.855 (2010) and 0.171 (2011). We can see that the number of interactions amongst specialists was decreasing, even after the INCT implementation in 2009. However, a positive difference (increase in

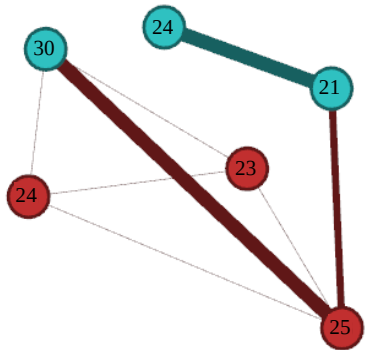

Figure 3. **Example for most connected nodes in a subnet.**

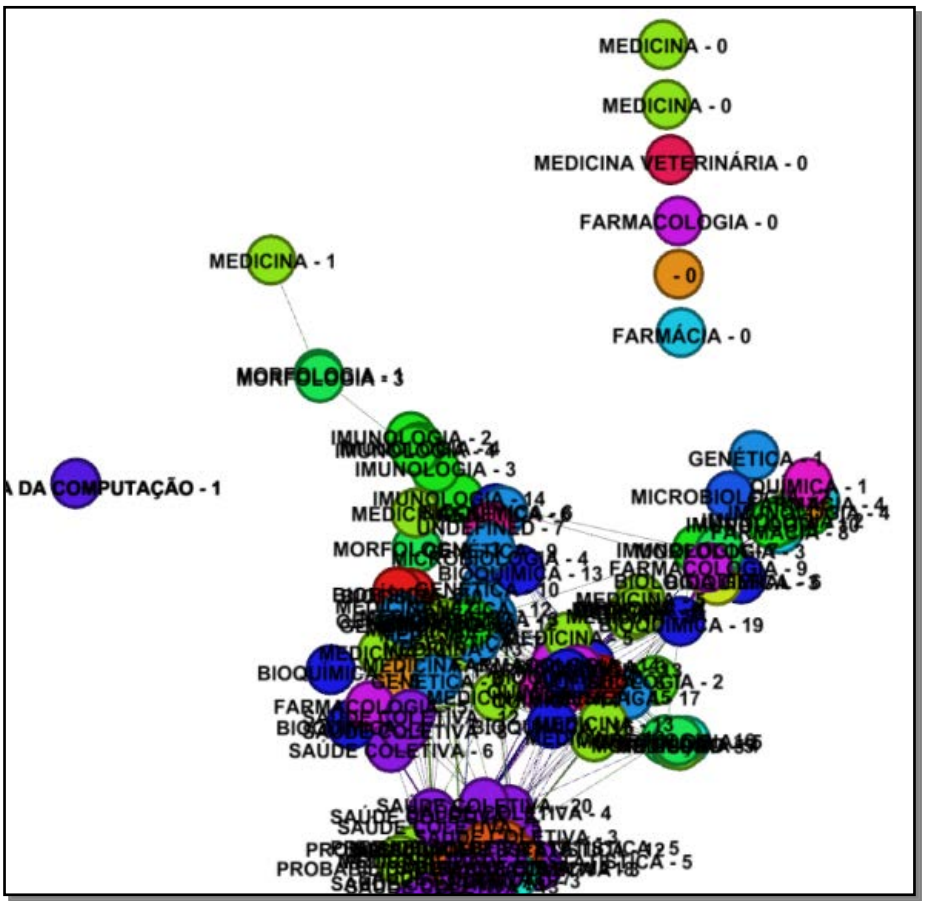

Figure 4. **Disconnected nodes.**

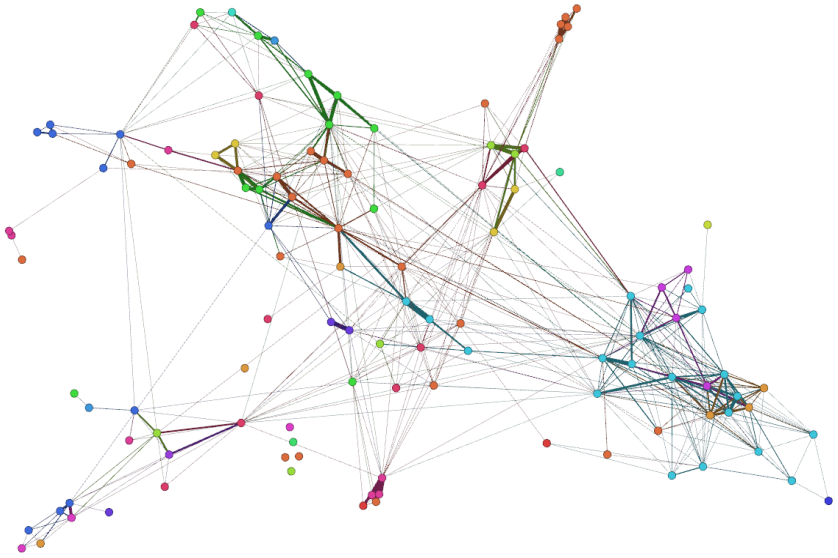

Figure 5. **INCTCC network.**

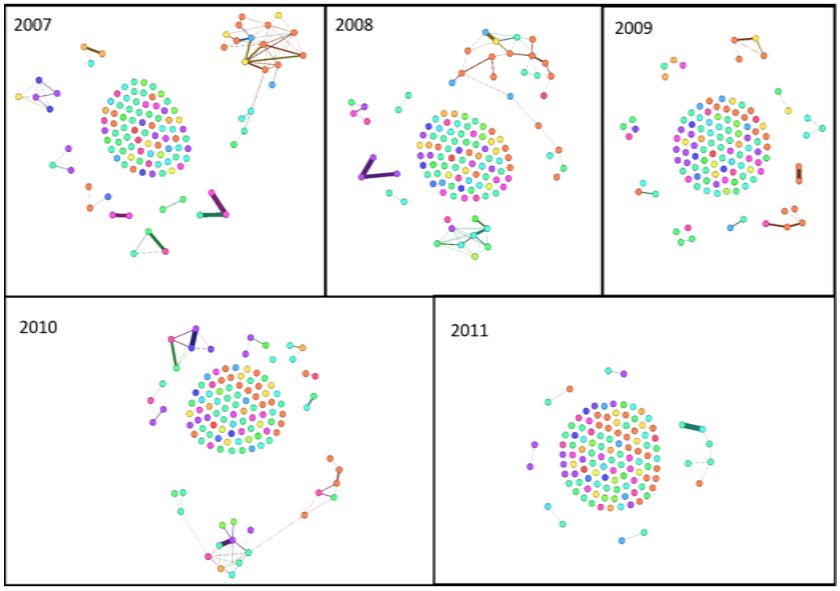

Figure 6. **INCTCC social network from 2007 to 2011.**

relationships) between years 2009-2010, the INCTCC's first year of operation, was of 0.445. It was higher than the difference of the previous year (2008-2009, a growth of −0.547), which was negative (decreasing of relationships), showing that interactions increased again after its creation. Meanwhile, in 2011 there was a significant decrease of this value. It possibly occurred as many publications were not yet registered with the Lattes Curriculum or were undergoing their review stage in the journals.

We can see it in **Figure 7**, which shows the total co-authorship interactions amongst researchers. There is an empowerment of relationships from 1993 until before the INCTCC's creation. This makes sense, as researchers knew each other and had constructed ties before the creation of the Institute.

Following the same line of reasoning, we saw that new relationships emerged from 2008 through to 2011, showing that the main goal of the INCTCC had being achieved. Year 2008 saw 54 new co-authorship relations, with 47 in 2009, 79 in 2010, and 25 in 2011. Again we should remember that when this data was processed, the production of 2011 was not 100% complete. Even with the decrease in the total number of relationships, the number of new relations remained almost constant, increasing in 2010.

We analyzed the number of publications over years, as shown in **Figure 8**. There is a 12% fall from 2008 to

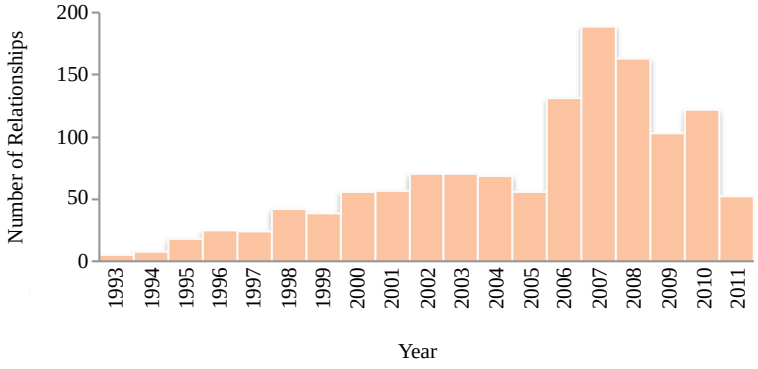

Figure 7. Total number of co-authorship relations over the years.

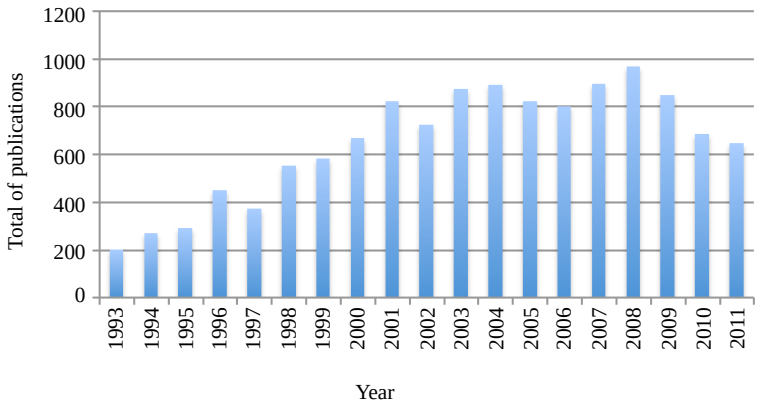

Figure 8. Total publications over the years.

2009, but the decrease in 2010, when compared to 2008 is of 29%. Probably after the INCTCC creation, researchers focused their research work on INCT areas. Another possible explanation is the increase in the new number of relationships. It is natural, when you start new professional interactions, that there is a period of adjustment. This adaptation process involves learning about one's new partner's works, understanding new processes and methods, and also the achievement of research maturity towards the obtaining of results. This adaptation consumes time, and fewer results are expected worthy of publication.

Figure 9 shows an example of interactions based on areas of common interest and expertise. The image shows interactions in the area of Collective Health. The edge's thickness indicates the number of interactions between two people.

With the developed environment, based on a multidimensional model, we can undertake several analyses. This project allows us to have a clear view of research group behaviour and identify key problems in scientific collaboration.

5. Related Work

The most similar work is from [7], whose focus is to analyze the social networks of researchers in the field of parasitic diseases such as dengue fever, Chagas disease and malaria, for example. Co-authorship was used to infer relationships amongst researchers in a particular area. Based on keywords, extracted from title of articles, the authors identified clusters. However, the difference to our study is the use of a higher number of datasets, as Lattes and PubMed, and the identification of different types of scientific relationships (not only co-authorship). Furthermore, we automated all the data treatment process. We also identify relationships amongst groups (not only amongst people), as example, institutions and funding agencies funding and can visualize all the interactions in a specific area.

The work presented by [16] tried to find patterns of interaction amongst researchers in the field of tourism in regions of Australia and New Zealand. For this, he used bibliometric information in the 1999-2005 period to analyze co-authorship networks, inter-institutional collaborations, and international collaborations. The similarity to our work is related to the use of metrics of social networks and some types of networks which were used. However, limitations in terms of viewing them compromise their final result. The fact that we created a visualization approach based on Gephi and Tableau helped us

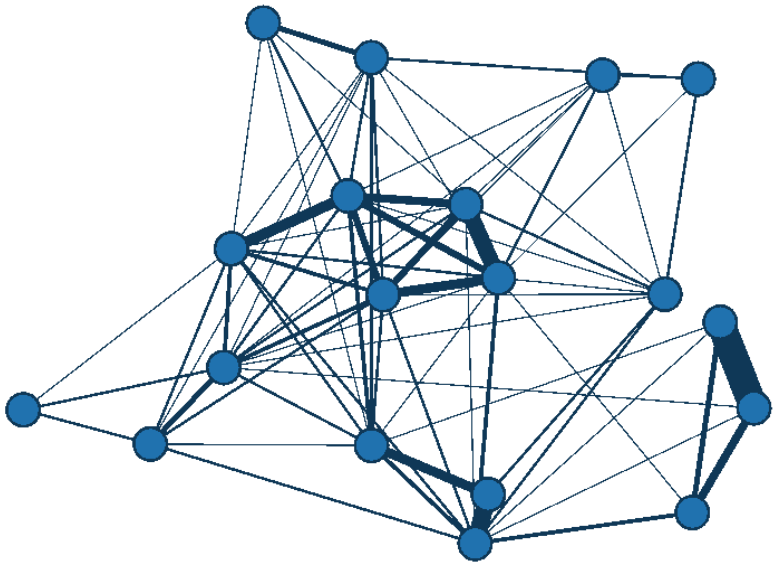

Figure 9. Interactions in the area of collective health.

significantly, as it provided more flexibility to configure metrics and parameters.

The research conducted by [17] is very close to ours, but their focus is on Web Science researchers. They do not use a multidimensional analysis to deal with data, and do not identify different kinds of relationships, either.

We can also mention the work of [18], who brought the concept of balancing and it was the foundation for the idea presented here. The main difference between the two work, Monclar's focused on a small community of the Department of Systems Engineering and Computing at COPPE/UFRJ, while this work focuses on all the Cancer Control INCT researchers. The Monclar *et al* work [2009] did not use a multidimensional database and therefore did not have the benefits of multidimensional analyses.

We also identified the work of [19]. It is a study on the behaviour of collaborative production, but focused on open-source software projects. The data used to identify relationships was obtained with the analysis of source code, discussion forums, chats, and version updates. The metrics were calculated to determine the structural characteristics (degree, centrality, etc.) and topological (density, diameter distribution, etc.) of social networks as well as in our work. The difference is that the focus was not to improve the network, but only to identify it.

In the study by [20], we see a method to detect, identify and visualize research groups in an university. The method is quite simple, relying on the generation of a matrix that lists the authors of the articles and it can infer a social network of co-authors. The visualization itself is quite clear and simple to understand, although it is not concerned with temporal analyses.

6. Conclusions and Future Work

Social network analysis helps to understand group development and to identify relationships patterns. This kind of analysis has been used in many situations, amongst which the health care scenario.

In this article, we presented the BRINCA project, whose goal is to support the analysis and visualization of scientific social networks. This project was applied in the area of Cancer Control in Brazil, in the scenario of the National Science and Technology Institutes (INCTs). The INCT is a mechanism to motivate collaboration amongst universities and research institutions dealing with strategic questions, in our case, cancer control.

This project is still under development and we can mention some future works and improvements. One of them is enriching the analysis, inputting data from medical records and cancer treatments. So, we can compare and identify the interaction amongst clinical treatment and research.

Another challenge is the adoption of data mining techniques to detect associative rules and recurrence patterns.

As last work, we will study the benefits of the approach created for the research scenario for cancer control in Brazil.

Acknowledgements

We would like to thank CNPq, CAPES, and FAPERJ for their support, specially by the support provided by the projects "INCT para Controle do Câncer" (CNPq 573806/2008-0 e FAPERJ E26/170.026/2008) and "Projeto Universal: CLOTO: Composição, Mineração, Análise e Predição de Redes Sociais Utilizando Dados Ligados Abertos e Contextualizado" (CNPq 487239/2012-1), by the

grants “Jovem Cientista do Nosso Estado” (Young Researcher of Rio de Janeiro, FAPERJ: E_23/2013) and “Produtividadeem Pesquisa-Nível 2” (Productivity in Research-Level 2, CNPq: 308219/2010-4).

REFERENCES

[1] INCTCC, “INCT Activity Report 2010—Home-INCA,” 2010. http://www1.inca.gov.br/inca/Arquivos/INCT/inct_project_2010.pdf

[2] R. A. Perorazio, F. F. Faria, R. Monclar, J. Oliveira and J. Souza, “Estudando Dinâmicas de Grupo Através da Utilização da Análise de Redes Sociais em uma Comunidade Médica,” *Proceedings of the Brazilian Workshop on Social Network Analysisand Mining*, *XXXII Congress of the Brazilian Computer Society*, Curitiba, 2012.

[3] J. C. Cordeiro, “Redes Sociais e Saúde,” *Revista Hispana para Elanálisis de Redes Sociales*, Vol. 12, No. 10, 2007.

[4] A. S. Klovdahl, “Social Networks and the Spread of Infectious Diseases: The AIDS Example,” *Social Science & Medicine*, Vol. 21, No. 11, 1985, pp. 1203-1216. http://dx.doi.org/10.1016/0277-9536(85)90269-2

[5] M. Negreiros, *et al*., “Optimization Models, Statistical and DSS Tools for Dengue Prevention and Combat,” *Efficient Decision Support Systems*: *Practice and Challenges in Biomedical Related Domain*, INTECH Opena Access Publisher, Vol. 1, 2011, pp. 115-160.

[6] R. S. Monclar, “Análise e Balanceamento de Redes Sociais no Contexto Científico,” M.Sc. Thesis, COPPE/PESC, Universidade Federal do Rio de Janeiro, 2008.

[7] C. M. Morel, S. J. Serruya, G. O. Penna and R. Guimaraes, “Co-Authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programmes on Neglected Diseases,” *PLoS Neglected Tropical Diseases*, Vol. 3, No. 8, 2009. http://dx.doi.org/10.1371/journal.pntd.0000501

[8] A. Parent, F. Bertrand, G. Côté, *et al*., “Scientometric Study on Collaboration between India and Canada, 1990-2001,” 2003. http://www.science-metrix.com/pdf/SM_2003_009_DFAIT_Indo-Canadian_S&T_Collaboration.pdf

[9] J. Owen-Smith, M. Riccaboni, F. Pammolli, *et al*., “A Comparison of U.S. and European University-Industry Relations in the Life Sciences,” *Management Science*, Vol. 48, No. 1, 2002, pp. 24-42. http://dx.doi.org/10.1287/mnsc.48.1.24.14275

[10] Lattes, “Lattes,” 2012. http://lattes.cnpq.br/

[11] Pubmed, “PubMed Home,” 2012. http://www.ncbi.nlm.nih.gov/pubmed

[12] Pentaho, “Pentaho Kettle Project,” 2012. http://kettle.pentaho.com

[13] Gephi, “Gephi,” 2012. http://gephi.org/

[14] Tableau, “Tableau Software,” 2012. http://www.tableausoftware.com/

[15] G. O. Fernandes, J. Oliveira and J. M. Souza, “XMLattes A Tool for Importing and Exporting Curricula Data,” *International Conference on Information and Knowledge Engineering*, Las Vegas, 2011.

[16] P. Benckendorff, “Exploring the Limits of Tourism Research Collaboration: A Social Network Analysis of Co-Authorship Patterns in Australian and New Zealand Tourism Research,” 20*th Annual CAUTHE Conference*, Hobart, Australia, 2010.

[17] A. H. F. Laender, *et al*., “Building a Research Social Network from an Individual Perspective,” *ACM/IEEE Joint Conference on Digital Libraries*, Ottawa, 2011, pp. 427-428.

[18] R. S. Monclar, J. Oliveira and J. M. Souza, “Analysis and Balancing of Social Network to Improve the Knowledge Flow on Multidisciplinary Teams,” 13*th International Conference on Computer Supported Cooperative Work in Design*, Santiago, Chile, 2009.

[19] S. F. De Sousa, M. A. Balieiro and C. R. B. de Souza, “Análise Multidimensional de Redes Sociais de Projetos de Software Livre,” *Proceedings of the* 2008 *Simpósio Brasileiro de Sistemas Colaborativos*, Vila Velha, 27-29 October 2008, pp. 23-33. http://dx.doi.org/10.1109/SBSC.2008.35

[20] A. Perianes-Rodriguez, C. Olmeda-Gómez and F. Moya-Anegón, “Detecting, Identifying and Visualizing Research Groups in Co-Authorship Networks,” *Scientometrics*, Vol. 82, No. 2, 2010, pp. 307-319. http://dx.doi.org/10.1007/s11192-009-0040-z