Meta-analysis of the TAEE project applying social network analysis

Jorge A. López Vargas, Nelson Piedra, Jeanneth Chicaiza

Computer Sciences School
Universidad Técnica Particular de Loja
Loja, Ecuador
{jalopez2, nopiedra, jachicaiza}@utpl.edu.ec

Edmundo Tovar

School of Computing
Universidad Politécnica de Madrid
Madrid, España
etovar@fi.ump.es

Manuel Blázques,Manuel Castro

Industrial Engineering School
Universidad Nacional de Educación a Distancia
Madrid, España
manuel.blazquez.merino@gmail.com,
mcastro@ieec.uned.es

Jesús Arriaga, Fernando Pescador

Control and Electronic Systems Department
Universidad Politécnica de Madrid
Madrid, España
{jesus.arriaga, fernando.pescador}@upm.es

Jesús Romo, Oskar Casquero, Manuel Benito, Ariana Landaluce

Universidad Del País Vasco
Bilbao, España
{jesus.romo, manuel.benito}@ehu.es,
ariana.landaluce@gmail.com

Abstract— The social network analysis (SNA) is an approach that can be applied as a complement to other analysis (such as statistical) in order to obtain other valuable information. The social network analysis has been used in several initiatives showing that it is an approach that can contribute in building the semantic web. Within the project Technologies Applied to Electronics Teaching (TAEE) there are biannual conferences (it has been organized since 1996) and have accumulated a significant amount of data resulting from the conferences held. All of this information constitutes a data source that should be exploited and that can provide meaningful information. In this document we describe, how to social network analysis has been used on data sources generated by user communities, in order to obtain some semantic artifacts, like ontologies. Also describes how to was applied the social network analysis and its metrics on the information generated in the TAEE congresses to answer a set of questions (What are the relationships and the level of cohesion of the different organizations (at the level of Spain and across continents) involved in TAEE? How have evolved the thematics covered in the conference?, What are the new ontological additions in technology over the years?, and How have evolved the thematics in the research and studies related to teaching electronics?) formulated by the organizers of the congresses and that through other approaches would have been a large task and complicated. The answers to the questions can provide us important information about the behavior and characteristics of the elements present in TAEE conferences, furthermore being an element for making decisions on future initiatives with the same style of TAEE.

Keywords- SNA, meta-analysis, TAEE, metrics, results, RedOBER

I. INTRODUCTION

The social network analysis (SNA) is an approach used to find not evident information within the structures formed by the interaction between user groups or entities. The less evident information is the basis for finding patterns of behavior that provide additional and relevant information about the operation and characteristics of a network of some kind. The Social networks analysis covers areas such as: conflict of interest, criminal's networks, finding influential individuals, and study of co-author's networks, among others.

Since a few years ago the social network analysis has been used in several initiatives that have selected as data source the content generated via Web 2.0 tools and have developed networks in order to apply concepts of social network analysis and find a starting point on the road to the Semantic Web, showing that the semantics can emerge from user communities and through simple actions such as tagging. Other initiatives have proposed expanding the Query Language for the Semantic Web (SPARQL) with some concepts and metrics of social network analysis to enable the discovery other types of relationships, derived in a context, which they have called semantic associations.

The project of Technologies Applied to the Teaching of Electronics (TAEE) is an initiative involving several educational organizations and companies from Europe, North America, Central America and South America. The aim of TAEE is implement massively new technologies to the teaching of electronics. Since 1996, a biannual conference has been organized by TAEE project that talks about technologies for teaching electronics and whose aims to disseminate results of projects and initiatives in the areas of educational innovation and application of new technologies in teaching electronics.

The articles of the TAEE's conferences have been classified into large ontological families such as: Systems, Devices and Components, Instrumentation and Measurement, Automation, Robotics, Laboratories, Educative Software, Teaching and Construction Techniques, Implementation and Practical Realization. The documentation generated by each article in

each conference is a data source to be exploited. To date there are; 964 documents, 1348 unique keywords, 1092 participating entities, and 1674 authors, among other data.

The aim of this paper is to show the results of the meta-analysis of the TAEE project's conferences (from 1996 to 2008) by applying social network analysis. Through some techniques, metrics and indicators of social network analysis was answered a set of questions that were asked by organizers of the congress and that are of your interest. Also was used to show some relationships such as the use of keywords in the description of the themes and the sub-structures that form between the co-authors of the documents. The metrics and indicators were used thus: to study the connections and distances between actors were used degree of a vertex and density, to determine the power and centrality of an organization were used degree centrality, closeness and betweenness, furthermore to analysis thematic and keywords, weight networks were used, Hubs, Authorities, and to study relations between co-authors k-cores was used.

The results obtained by applying social network analysis on the congresses of TAEE project unlike other analysis (such as statistical), has permitted us to: see graphically the relations between the different participants as well as the development of ontological families. Likewise, the task of finding isolated organisms and those organisms that serve as a link or bridge with others went simple, and showed the way in which they have been used keywords in different ontological families.

The paper is structured as follows: in the first section, we will discuss the social network analysis, its importance and its implementation in some projects of Semantic Web. The next section will be devoted to the TAEE Project's data source for meta-analysis. The third section will show the application of social network analysis and the metrics used to answer the questions described above. Finally, a section of conclusions and recommendations.

II. SOCIAL NETWORK ANALYSIS

A social network is a social structure made of individuals (or organizations) called "nodes," which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige[1]. As long as, as mentioned in [5], the social network analysis offers the methodology to analyze social relations; it tells us how to conceptualize social networks and how to analyze them.

The social network analysis (SNA) is an approach used to find information not evident within the structures formed by the interaction between user groups or entities. As you can read in [5] the objective is to detect and interpret patterns, in the social ties between actors that provide additional relevant information about the operation and characteristics of a network of some kind.

Because of its inherent power to reduce a system to its individual components and their relationships (network

characterization), moreover the existence of several metrics (sociometric) and indicators that characterize these structures, is an approach that has gained popularity in recent years, successfully applied in scientific and technological disciplines as diverse, mentioned in [13], some examples of projects: conflicts of interest [1], criminal networks [14], search of influential individuals [15], study of co-authors networks [3], among others.

For a few years ago the social network analysis has been used in several initiatives, such as [12], which have selected as a data source the content generated via Web 2.0 tools and have developed networks in order to apply concepts of social network analysis and find a starting point on the road to the Semantic Web, showing that the semantics can emerge from a users communities and through simple actions such as tagging. For the creation of ontologies, the social network analysis has been used on data repositories created by user's communities to find lightweight ontologies, lacking some formalities such as restrictions, but they cover a large number of entities that have been created to through simple actions, but extremely participative.

Within this issue is important to mention the model of ontologies, Actor-Concept-Instance proposed in [12] inspired by social tagging mechanisms, creates semantic-social networks represented in a tripartite graph, person, concepts and instances, expanding the traditional concept of ontology (concepts and instances) with a social dimension.

According to the model, a network of folksonomies is a hypergraph with ternary sides, where each side represents the fact that an actor (user) associated with a certain instance (resource) with certain concepts (tags). From the ternary hypergraph, is possible obtain three bipartite graphs (actors and concepts, concepts and instances, and actors and instances) of whom (actors, concepts) shows an ontology formed by the interests of communities, while (concepts, instances) reflect ontologies formed by the co-occurrence of tags in the resources, thus showing how from a folksonomy can be deduced an ontology.

In SPARQL, arises the need to incorporate, to RDF graphs, some concepts and metrics own of social network analysis, such as: length between nodes, the search for shorter paths, among others, and this way enlarge the information than a SPARQL query provides, as show [7], [1], [14], [8], [2]. With these information extensions, SPARQL will be able to find certain relationships that have been called semantic associations or complex semantic relationships. Semantic associations are meaningful and relevant complex relationships between entities, events and concepts. They lend meaning to information, making it understandable and actionable, and provide new and possibly unexpected insights [14]. Discovery of semantic associations is the process of finding paths of possibly unknown length that connect the given entities and have a specific semantics [8].

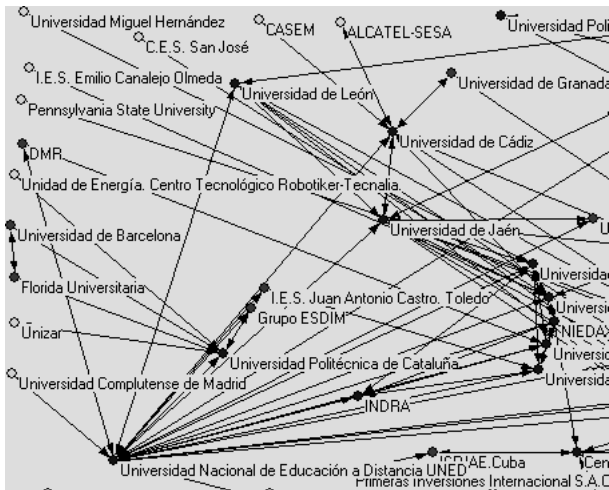1 Taken from Wikipedia http://en.wikipedia.org/wiki/Social_network

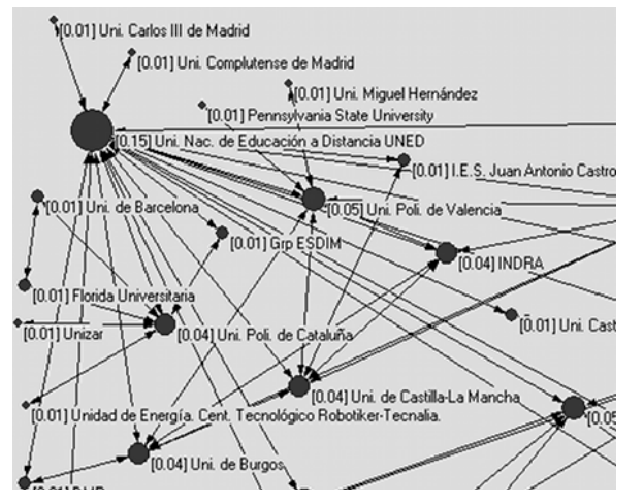Figure 1. A section of the network constructed

Figure 2. Vertex degree. Each vertex shows your degree

The SPARQL extensions are potentially applicable to: discovery of conflicts of interest, the study of criminal networks, for example.

Other applications of social network analysis, similar to the analysis described in this document can be found in [9], [10], [11] and [13] whom have taken as a data source code repositories of open source projects, including: Apache, GNOME and KDE; building a methodology which highlight the relevant information for a given aspect of project, as mentioned [10].

Another similar application can be found in [6] in this case the data source are communication archives, such as e-mails, blogs, with the aim of improving the understanding of organizational behavior in a company.

In [4] was used social network analysis to detect interesting structures in the resource repository called SIO (Slovenian Educational Network).

III. TAEE PROJECT AS DATA SOURCE FOR THE META-ANALYSIS

Under the project of Technologies Applied to the Teaching of Electronics[2] since 1996, a biannual conference has been organized that talks about technologies for teaching electronics and whose aims to disseminate results of projects and initiatives in the areas of educational innovation and application of new technologies in teaching electronics. The presentations prepared for each conferences are varied but have been classified into large families themes such as: Systems, Devices and Components, Instrumentation & Measurement, Automation, Robotics, Laboratories, Educative Software, Teaching and Construction Techniques, Implementation and Practical Realization. The documentation generated by each article in each conference is a source of data to exploit and seize.

For the development of this article we took as a data source, the documents generated in the congress of TAEE Project, 8 congresses executed (from 1996 to 2008) organized in 150 sessions, which has resulted, in a data group consisting of: 964 documents, 1348 unique keywords, 1092 participating agencies, 1674 authors, 9 thematic families, several levels of specialization in each of the families, 4425 learning objects extracted from each article and approximately 5000 references.

The data listed above were provided by the organization of TAEE congresses and were delivered in two formats, spreadsheet and a database of MS Access.

To perform the meta-analysis was used Pajek[3], an open source tool designed for representation of large volumes of data. There are other tools like Pajek, how can see in "Social Network Analysis: Introduction and Resources"[4]
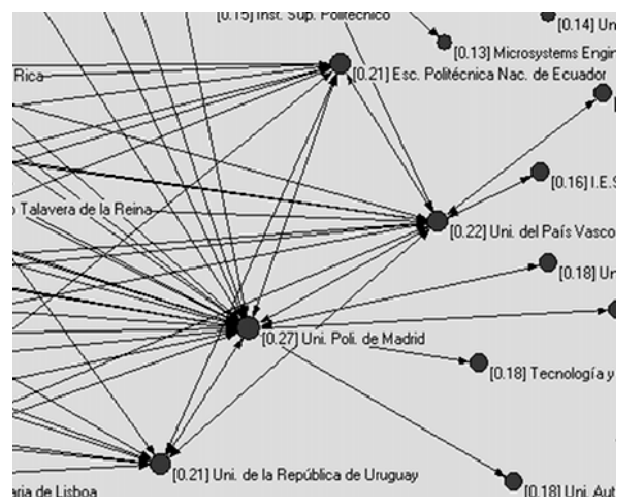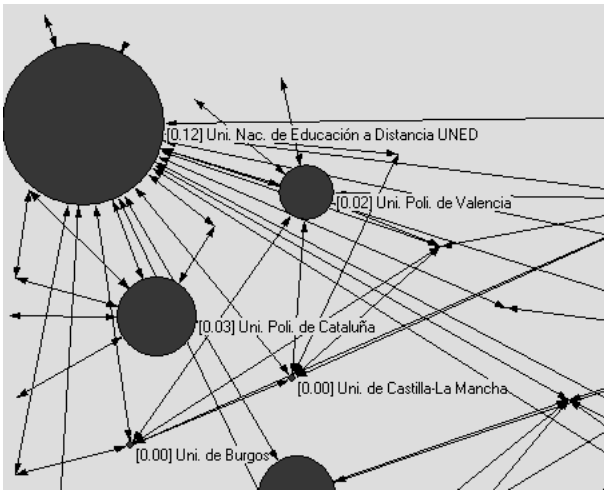

Figure 3. Closeness

Figure 4.   Betweeness

Most of the data for meta-analysis were obtained through SQL queries executed against the database and the results were exported as plain text files. In some cases Java programs were developed that allowed to obtain structured data files in the format of Pajek.

IV.   APPLICATION OF SOCIAL NETWORK ANALYSIS

Through meta-analysis, we have attempted to answer the questions: What are the relationships and the level of cohesion of the different organizations (at the level of Spain and across continents) involved in TAEE?, How have evolutioned the thematics covered in the conference?, What are the new ontological additions in technology over the years?, and How have evolutioned the thematics in the research and studies related to teaching electronics?

To answer these questions, we have used several metrics of social network analysis, such as vertex degree and density (to study connections and distances between the actors), degree centrality, closeness and betweenness (to determine those organizations with greater power and centrality). Furthermore, we have used Weighted networks, Hubs and Authorities for analysis of thematic and keywords. Lastly K-cores for the study of substructures formed by the relationships between the authors of documents. The results obtained by applying social network analysis on the Congress documents of TAEE project is a complement to statistical analysis, and has permitted: see graphically the relations between the different participants as well as the development of thematic families by facilitating the task of finding isolated organisms and those organisms that serve as a link or bridge with others, and has identified the relationship between keywords in different thematic families.

TABLE I.   SUMMARY OF THE VERTEX DEGREES

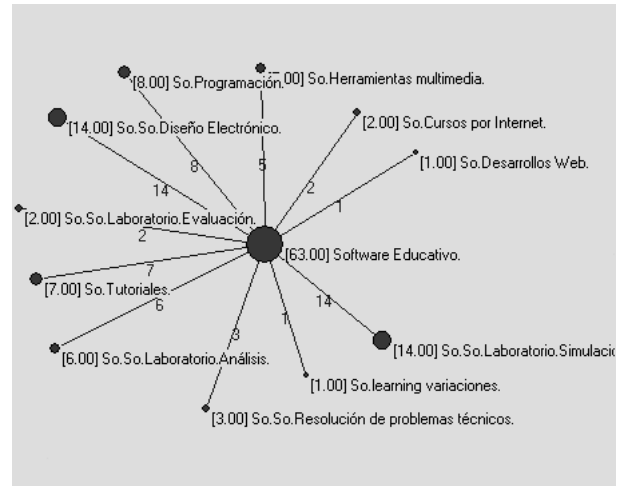| Degree	Frequency	Representative
0	51	SELCO. SA
2	36	ALCATEL-SESA
4	13	DMR
6	8	AYDE, S.L.
14	6	I.E.S. J.A. de Castro Talavera de la Reina
20	6	Cent. Fed. de Educacao Tecno do Paraná. Brasil
42	1	Universidad Nacional de Educación a Distancia UNED
52	1	Universidad Politécnica de Madrid UPM

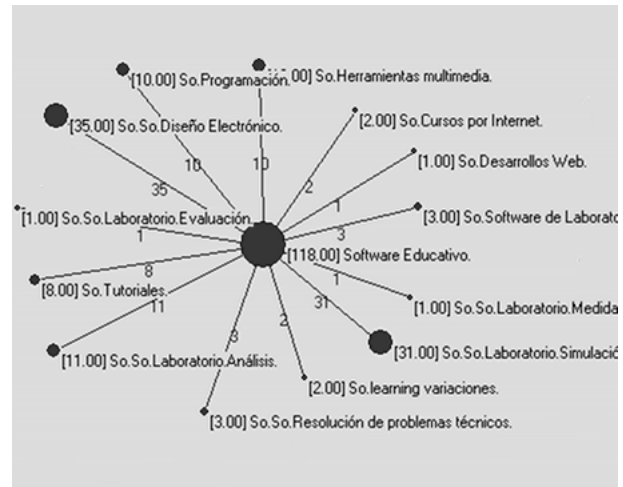Figure 5.   Thematic classification to Educative Software in 1994

Figure 6.   Thematic classification to Educative Software in 1996

A.   What are the relationships and the level of cohesion of the different organizations (at the level of Spain and across continents) involved in TAEE?

To answer this question was built a network composed as follows: the nodes or actors representing the organizations participating in any edition of the congress. The arcs of the graph are formed between organizations that cooperated in the creation of any presentation, at least once; due to lack of hierarchies, the arcs are bidirectional. See Fig. 1.

The indicators and metrics used to answer the question were: cohesion, density, centrality, vertex degree, closeness, and betweenness. Before to continue is necessary some definitions.

Cohesion is the degree to which actors are connected directly to each other, forming groups, and intuitively means that a social network contains many arcs [5]. To determine the degree of cohesion there are some metrics that are described briefly below:

Teaching	1994	1996	1998	2000	2002	2004	2006	2008	Año
Análisis Si. enseñanza experiencias didácticas	X	X	X	X	X	X	X	X	1994
Aprendizaje basado en tareas	X	X	X		X	X	X	X	1994
Aprendizaje basado en proyectos	X	X	X	X	X	X	X	X	1994
Metodología docente	X	X	X	X	X	X	X	X	1994
Estudios relacionados con educación	X	X	X	X	X	X	X	X	1994
Espacio Europeo Educación Superior						X	X	X	2004
Docencia a distancia		X		X	X	X	X	X	1996
Ap. On line	X	X	X	X	X	X	X	X	1994
Material educativo y docente	X	X	X	X	X	X	X	X	1994
Innovación docente	X			X	X	X	X	X	1994
Aprendizaje basado en prácticas	X	X	X	X	X	X	X	X	1994
Mode									1994
Frequency and median	9	9	8	9	10	11	11	11	M_e = 9.5

- Density: is the proportion of existing ties or arcs in a network with relative to the total number of possible arcs. The density calculated for the network of all organizations that participated in a conference is 0.0191 which means that only 1.91% of all possible arcs are present, although as noted in [5] the density is not a useful measure of their dependence on network size. See Fig. 2.

- Centrality: This measure provides an approximation of the social power of a node on the basis of how well "connected" found on the network. For the calculation of centrality, the following indicators: degree, closeness and betweeness have been used.

  o Degree: The vertex degree is a count of the number of ties or arcs (incoming, outgoing or both) with other actors or nodes in the network, for summary sees Table 1.

  o Closeness: It is the measure that has a node and shows how close is (directly or indirectly) to all other nodes in a network. It reflects the ability to access information through the group of network members. It is inversely proportional to the sum of the shortest distance between all nodes in the network. See Fig. 3.

  o Betweeness: This measure shows how a node is, compared to other network nodes. This measure looks at the connectivity of the neighboring node, assigning a higher value for the nodes that serve as bridges between clusters. The measure reflects the number of nodes that a node connects indirectly through their direct links. See Fig. 4.

| Weighted arcs			Hubs y Authorities		
Order	Keyword	Weigth	Order	Keyword	Weigth
1	instrumentación	174	1	EEES	0.04
2	PCB	192	2	multimedia	0.04
3	multimedia	200	3	laboratorio	0.06
4	docencia	286	4	docencia	0.07
5	laboratorio	315	5	diseño	0.08
6	análisis	326	6	análisis	0.09
7	diseño	367	7	simulación	0.09
8	metodología	409	8	metodología	0.10
9	simulación	420	9	práctica	0.11
10	práctica	550	-	-	-
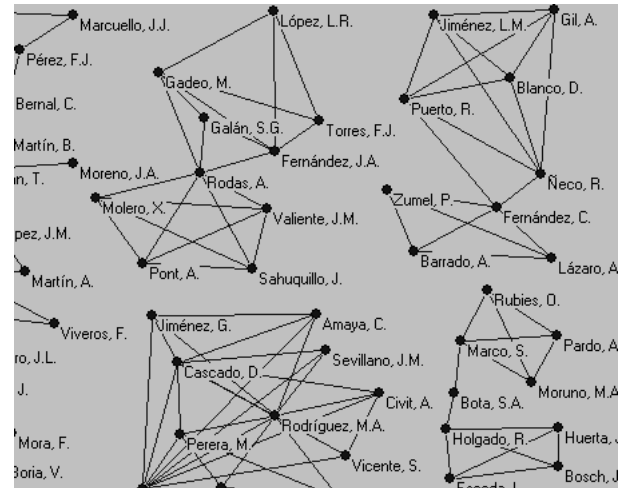
Figure 7. Part of k-cores in the authors network

B. How have evolutioned the thematics covered in the conference?

To answer the question several networks were formed, one by year of publication of the congress, whose nodes are the thematic (which remained fixed in all the years) and their sub-thematic (which appear and disappear depending on the existence of documents classified under this sub-thematic). Giving as result 9 groups, one for each thematic, around which are shown, according to existence of documents, the sub-thematic.

For the arcs, we used weighted arcs, the weight representing the number of documents belonging to a sub-thematic. The diameter of the nodes varies proportionally to the number of documents that were published, it should be noted that the nodes that represent the thematic, show the total amount of documents.

As mentioned above the metric used was the weighted arcs, which is a value assigned to an arc to represent a certain characteristic and the weights can be classified into input, output or both.

In the following figures (Fig. 5. and Fig. 6.) you can see part of the evolution of the Educative Software thematic in the events of the years 1994 and 1996. To the generation of images we used an option of Pajek called "Generate in Time" [16] option that generates a series of images from a single data file.

April 14-16, 2010, Madrid, SPAIN

Through the series of generated images, was possible: classify into groups according to the level of specialization; get the years in which different thematic achieved its minimum and maximum levels of specialization, and find the families whose thematic had a lower index changes.

C. What are the new ontological additions in technology over the years?

To answer this question were selected the following thematic: Systems, Educative Software, Devices and Components, Automation and Robotics, considering them, that are the families more related to the area of technology. In addition, use the metric called the weighted arcs and the statistical measures: mode and median. The mode, allow get the common year in which they appeared each sub-thematic, while the median is used to determine the number of sub-thematic.

The above has helped focus attention on a specific subset and provides numerical arguments that describe the group's situation. Demonstrating that social network analysis needs to be complemented with statistics metrics and vice versa.

D. How have evolutioned the thematics in the research and studies related to teaching electronics?

Like the previous question was used: the weighted arcs, mode and median to answer this question, but the thematic family selected was Teaching. The following table, Table 2 shows a summary of the data source to calculate the mode and median.

E. Another analysis

The following analysis not responds to a specific question, but we use it is to show relevant information about the network behavior of thematic families and keywords, and the network of authors and co-authors of articles.

- Thematic families and keywords. This analysis tries to show the use of keywords in the description of the documents that belong to a thematic family. It formed a network of nodes representing the thematic families and keywords, there is a tie between thematic family and keyword if there are one or more documents that belongs to a thematic family and was described with that keyword. To perform the analysis, the following metrics were used: arcs weighted, Hubs and Authorities, this metrics try to show that a keyword is important not only for the number of times it is used, but instead it is important for the way in which used. Hubs and Authorities are useful metrics for finding important vertices inside a network structure. A node is a good Hub if it points to a number of vertices that are good Authorities, and a node is a good Authority, if it is tied to several vertices, that are good Hubs. The most important arcs both by the weighted arcs, and by the Authorities analysis are: Teaching, Educative Software and Systems, while keywords (Hubs) are summarized in the following table, see Table 3.

In the analysis of Hubs and Authorities, and as shown in Table III, only 9 keywords are showed, because the Pajek did not generate the tenth, possibly due to the lack of a keyword that is used to describe the 3 Authorities. We can see how the ranking of keywords varies depending on the type of analysis, even; we can see the omission or addition of other.

- Authors and co-authors. The objective of this analysis was to show the groups constructed by the authors of published documents, and beyond the simple groups, we can see how simple groups interact to form larger groups. To do this analysis, we formed a network whose nodes are authors and there is an arc between two nodes, if two authors have collaborated to write an article. The metric used was K-cores. A k-core is a maximal sub-network[5] in which each vertex has at least degree k within the sub-network [5]. It uses the degree of a vertex to identify groups (clusters) of vertices that are strongly connected because each vertex has the same level within the group. The value of index k was selected to display as many authors as was possible. Fig. 7. show part of the graph. In the case of the Fig.7. we set the value of k to 4, which means that groups are formed by nodes that are related to 4 nodes, ie the degree of each node is equal to 4.

V. CONCLUSION

Through the development of this work, we can draw the following conclusions:

- The answers to the questions raised were developed using techniques of social network analysis. Demonstrating that most activities that involve a group of people can be represented through a graph and can apply techniques of social network analysis to characterize them.

- The social network analysis is a type of analysis that allows to obtain other relevant information, complementary to other analysis like statistical and, through the metrics and indicators that it has, is possible to show with numerical data, the different characteristics that a social networking has.

- Through analysis of the relationships between thematic families and keywords, we can see how social network analysis provides a different approach that shows other information in a simple, fast and graphically.

- The meta-analysis just presented in this document has served to provide insight from massive data analysis and statistics. From this point of view has been included joint variables parameterization based social networks and graph theory by using the Pajek tool.

- TAEE congresses have followed a positive development reflecting the electronics world in terms

---

5

A maximal subnertwork is a network with maximum density

of thematic diversity and their adaptation to new techniques and products.

- However, the partnership between organizations and between authors from different organizations is a topic that could be improved, in addition to achieving a better balance between organizations from different continents. From this paper, the authors support the dedication of all the organizing committees of various conferences and summons them to encourage to new organizations to participate and strengthen partnerships between organizations.

In order to generalize this type of analysis for any repository generated through user communities, we make the following recommendations, attempting to mitigate some risks described in [12]:

- The ternary hypergraph, could be formed by authors (actors) who write documents (resources) that are described by keywords (tags).

- To achieve the volume of information necessary to ensure the accuracy of the generated ontologies, is necessary the publication of documents on a public website, accessible through the use of Web 2.0 tools that allow the enrichment level of tags, for the resource description.

- To avoid some of the problems in creating the folksonomy, especially to avoid the use of nonstandard acronyms and errors in writing, we propose the use a suggestion service (but not with the intention to limit, restrict or impose criteria to users) without losing the social participation and freedom to choice.

REFERENCES

[1] Boanerges Aleman-Meza, Meenakshi Nagarajan, Li Ding, Amit Sheth, I. Budak Arpinar, Anupam Joshi, and Tim Finin. Scalable semántica analytics on social networks for addressing the problem of conflict of interest detection. *ACM Trans. Web*, 2(1):1–29, 2008.

[2] Kemafor Anyanwu, Angela Maduko, and Amit P. Sheth. Sparq2l: towards support for subgraph extraction queries in rdf databases. In *WWW*, pages 797–806, 2007.

[3] Albert-László Barabási. Linked: The new science of networks. J. *Artificial Societies and Social Simulation*, 6(2), 2003.

[4] Vladimir Batagelj, Iztok Kavkler, and Matija Lokar. Network análisis of repositories. In LODE, 2007.

[5] Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj. *Exploratory Network Analysis with Pajek*. Cambridge University Press, 2005.

[6] P.A. Gloor and Yan Zhao. Analyzing actors and their discussion topics by semantic social network analysis. pages 130–135, July 2006.

[7] Chris Halaschek, Boanerges Aleman-meza, I. Budak Arpinar, and Amit P. Sheth. Discovering and ranking semantic associations over a large rdf metabase. *In In 30th International Conference on Very Large Data Bases*, 2004.

[8] Krys Kochut and Maciej Janik. Sparqler: Extended sparql for semantic association discovery. In *ESWC*, pages 145–159, 2007.

[9] Luis Lopez-Fernandez, González Barahona, and Gregorio Robles. Applying social network analysis to the information in cvs repositories. In *Proceedings of the Mining Software Repositories Workshop. 26th International Conference on Software Engineering*, 2004.

[10] Luis Lopez-Fernandez, Gregorio Robles, Jesus Gonzalez-Barahona, and Israel Herraiz. Applying social network analysis techniques to community-driven libre software projects. *International Journal of Information Technology and Web Engineering*, 1(3): 27–48, 2006.

[11] Juan Martinez-Romo, Gregorio Robles, Jesús M. González-Barahona, and Miguel Ortuño-Perez. Using social network analysis techniques to study collaboration between a floss community and a company. In *OSS*, pages 171–186, 2008.

[12] Peter Mika. Ontologies are us: A unified model of social networks and semantics. *Web Semant.*, 5(1):5–15, 2007.

[13] Gregorio Robles. *Empirical Software Engineering Research On Libre Software: Data Sources, Methodologies and Results*. PhD thesis, Escuela Superior de Ciencias Experimentales y Tecnología - Universidad Rey Juan Carlos, Móstoles, Madrid, Spain, 2005.

[14] A. Sheth, B. Aleman-Meza, F. S. Arpinar, Amit Sheth, C. Ramakrishnan, C. Bertram, Y. Warke, K. Anyanwu, Boanerges Aleman-meza, I. Budak Arpinar, , K. Kochut, C. Halaschek, Cartic Ramakrishnan, Yashodhan Warke, D. Avant, F. Sena Arpinar, Kemafor Anyanwu, and Krys Kochut. Semantic association identification and knowledge discovery for nacional security applications. *Journal of Database Management*, 16:33–53, 2005.

[15] Alan F. Smeaton, Gary Keogh, Cathal Gurrin, Kieran McDonald, and Tom Sodring. Analysis of papers from twenty-five years of sigir conferences: what have we been doing for the last quarter of a century? *SIGIR Forum*, 37(1):49–53, 2003.

[16] Batagelj Vladimir and Mrvar Andrej. *Pajek Program for Analysis and Visualization of Large Networks*, Dec. 2008.

April 14-16, 2010, Madrid, SPAIN
IEEE EDUCON Education Engineering 2010 – The Future of Global Learning Engineering Education

135