IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 29, 2020, accepted July 14, 2020, date of publication July 27, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011947

Characterization of the Chilean Public Procurement Ecosystem Using Social Network Analysis

FABIOLA HERRERA LEIVA[1], ROMINA TORRES[1], (Member, IEEE), ORIETTA NICOLIS[1], AND RODRIGO SALAS F[2], (Senior Member, IEEE)

[1]Facultad de Ingeniería, Universidad Andres Bello, Viña del Mar 2531015, Chile
[2]Escuela de Ingeniería Civil Biomédica, Universidad de Valparaíso, Valparaíso 2362905, Chile

Corresponding author: Romina Torres (romina.torres@unab.cl)

This work was supported in part by the Universidad Andres Bello Internal Project under Grant DI-02-19/R.

ABSTRACT "Mercado Público" is a Chilean electronic platform used for purchasing processes by Chilean public organizations for the last two decades. The main aim of this study is to characterize the Chilean public procurement ecosystem by using social network analysis to detect the main communities of suppliers based on who awarded the tenders. To do this, we use a methodology that first represents the bidder-supplier relationship as a bipartite graph using purchase order information. Then we project the bipartite graph onto a monopartite graph of suppliers. We end by detecting the main supplier communities using a modularity algorithm. When we applied this methodology to the large tender segment in the Chilean public procurement market over a period of four years, we successfully detected the five largest communities and the micro and small companies which had the greatest rate of participation over time.

INDEX TERMS Bipartite graph, communities, modularity, public procurement, tenders.

I. INTRODUCTION

Mercado Público is the public procurement system of the Chilean State and the largest electronic trading platform in Chile, through which most public goods and services are traded. In this system it is possible to publish, store, distribute and analyze information about purchases made by the main offices of the State. The reason for centralizing transactions in "Mercado Público" is to promote transparency and efficiency in the contracting process between state purchasers and suppliers. Annually, more than 10 billion dollars (about 3.5% of Chile's Gross Domestic Product) are traded on the platform, which lists 123,000 suppliers.

The operation of "Mercado Público" starts with the publication of a tender on the platform www.mercadopublico.cl, a procedure carried out independently by buyer agency where interested suppliers are invited to submit bids to provide goods or services in order for the buyer agency to select the most convenient offer according to the criteria established in the bidding rules. These rules establish the requirements, conditions and specifications of the product(s) or service(s) to be contracted, the evaluation criteria that will be applied in the process, and the associated guarantees of each good or service. The tenders are awarded to those vendors that have made the best offers according to those evaluation criteria. It is important to note that a tender may be awarded to more than one supplier. Each supplier's participation percentage in the awarded tender is defined in the purchase order, and they do not necessarily need to have a prior agreement among themselves.

Table 1 shows the different types of tenders, where large tenders are those greater than 1,000 Monthly Monetary Unit (UTM) where each UTM had a value of USD $ 64.7 at February 2020. Table 2 classifies Chilean companies according to their annual sales. During 2016, 93% of suppliers that did business with the State were micro and small companies and their participation in sales reached 45%. While these results were encouraging, Chile hopes to increase the participation rate of micro and small companies, especially in the area of large tenders. This is an international challenge. For instance, on the one hand, in the United Kingdom heritage sector, studies have been carried out to understand the behavior

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

TABLE 1. **Types of Tenders.**

Type	Description
E2	Private tenders under 100 UTM
CO	Private tenders in a range between 100 to 1,000 UTM.
B2	Private tenders in the range of 1,000 to 2,000 UTM.
H2	Private tenders in a range between 2,000 and 5,000 UTM.
I2	Private tenders equal to or greater than 5,000 UTM.
L1	Public tenders under 100 UTM.
LE	Public tenders in the range of 100 to 2,000 UTM.
LP	Public tenders in a range between 1,000 and 2,000 UTM.
LQ	Public tenders in a range between 2,000 and 5,000 UTM.
LR	Tenders equal to or greater than 5,000 UTM.

TABLE 2. **Suppliers Classification according to the Company Size.**

Company size	Annual Sales
Large	Over four million USD
Median	Under than four million but over one Million of USD
Small	Less than one million but over 100,000 USD
Micro	Less than 100,000 USD

and successful participation of small and medium companies, where different strategies are recommended for each company according to its size [20]. Other countries have encouraged the state to do the business (public procurement) with minority-owned businesses through legislation [5].

The aim of this paper is to understand and characterize the Chilean public procurement ecosystem. The primary objective is to reconstruct the public procurement system by applying social network analysis to purchase order information in order to detect the main suppliers' communities and compare them over time. The analysis is performed by representing the interaction of tenders and suppliers using graph theory. Specifically, the ''Mercado Público'' is represented through a main graph composed of several interrelated bipartite graphs. The characterization of the graph enables information to be obtained from the network that allows the identification of suppliers' communities in the procurement market.

As far as we know, there are no other studies that characterize the public procurement market in terms of its main communities using social network analysis with graph theory, the proposed methodology is therefore a novelty in this field. Similar approaches have been proposed by the following works. A general study of public expenditure in Greece using graph analysis [29] (which does not detect communities), where public expenditure is represented as a network where nodes (public entities and beneficiaries) are connected by payments. A second similar work was performed in electronic commerce, where graph theory is also used but where the detected communities are the customers that bought products [15]. Another work recommended products of interest to customers based on information about what other products had been bought by customers of the same community [27]. Other authors [2] built a network to establish proper co-marketing strategies, where different communities of consumers were discovered by analysis of consumption patterns.

This paper is organized as follows: Section II, is a bibliographic discussion on existing literature about the detection of communities with graphs. In Section III, we present a methodology to conduct the study of the Chilean public procurement system by using social network analysis. The results are presented in Section IV. Conclusions and suggestions for future study are given in Section V.

II. BACKGROUND

The study of complex systems represented through graphs has seen important developments in the last few years [10]. Real networks that are not homogeneous and present a high level of order and organization may be represented using graphs. Several low-grade patterns coexist with high-grade vertices and the distribution of links is not only global, but also locally heterogeneous, with high concentrations of links in special groups of nodes and low concentrations among those groups. This characteristic of real networks is called community. A community is a cluster whose analysis is of great importance in the characterization of networks, since the groups of nodes probably share common properties or similar roles or functions within the network [10]. Typically, networks are represented as monopartite networks, where the graphs' nodes are all of the same type. Bipartite networks are a very important and common type of complex network, and many real-world networks are naturally bipartite [30], for instance, the network of actors and films [21], the scientific collaboration network [22], the artistic collaboration network [12], to name a few. In general, community detection techniques have been applied in different areas of knowledge, such as sociology, biology, information technology, and astronomy, and may have specific applications in social networks [14], communications networks [4], and grouping web clients with geographically similar interests to improve performance of dedicated web servers [17].

Depending on the size of the network, the characterization and interpretation of the clusters in the communities is done in different ways. On the one hand, networks with a small number of nodes and links can be described through a visual inspection of the communities detected, based on the attributes of their nodes and links. On the other hand, in the case of larger networks, algorithms are very helpful in identifying those attributes of nodes and links that occur most often within the community. In those cases, whose descriptive attributes contain non-categorized text-type values, techniques are used to debug data and filter out relevant terms to obtain meaningful elements that objects can represent.

Fortunato has reviewed and classified the community detection methods proposed in the last few years, separating them by the type of algorithm used [10]. This study shows that the selection of the algorithm depended on the type, topology and quantity of vertices and links in the network.

Graph partitioning is one of the most well-known traditional methods, where the network represented as a graph is partitioned in a pre-specified number of clusters. This is a disadvantage for the purpose of community detection. When it is necessary to understand the community structure of a graph, hierarchical clustering is useful.

Modularity algorithms are also a useful method to identify communities in a graph. In a bipartite network there are two disjointed sets of nodes, called upper nodes and lower nodes. The links connect only a pair of vertices that join both sets. Girvan and Newman [11], Newman [24], Newman and Girvan [23] proposed an algorithm to identify communities based on the modularity measure (defined as the fraction of links within a community in the network minus their expected fraction of links in a random network). Due to researchers having claimed that the original expression of modularity is not ideal for detecting communities in bipartite graphs, they have proposed first to project the graph onto a monopartite type network [10] in order to apply community detection techniques on this new network. According to Fortunato [10], this is useful in large networks and when the number of communities to be detected is not known a priori.

In the case of large networks with millions of links and nodes, Blondel *et al.* [4] proposed a simple method to extract the structure of the community. The heuristic method is an optimization of the greedy method and it is based on the optimization of modularity. It was demonstrated that this new method outperforms all other known methods of community detection in terms of time. In addition, the quality of the community is very good, measured by the value of modularity of the network, validated by an experiment applied to a mobile telephone network of two million customers and also a second experiment analyzing a web graph of 118 million nodes and more than a billion links [4].

When detecting communities in real networks, nodes can belong to more than one community at a time. Palla *et al* [26], developed a technique to find overlapping communities by testing their effectiveness in three different applications (network of co-authors of scientific articles, protein network and network of word communities), in order to interpret the global organization of the network from the coexistence of the communities, to understand the structural and functional properties of the networks.

Lately, new techniques to speed up current algorithms have been proposed [19] as well as different techniques for studying evolution over time once the communities have been detected (e.g [1]).

Once the communities are detected, there is an area of research that is devoted to analyzing their evolution in different periods of time, which has been addressed by different authors. According to Fortunato [10] problems arise when attempting the detection of dynamic communities, efforts being concentrated on the static version of the problem. From this perspective, the main phenomena of community life cycles are birth, contraction, fusion with other communities, separation, and death. Thus, it becomes important to follow the evolution of the structure of the community over time (which allows us to discover how communities are generated and how they interact with each other). Some authors have analyzed the evolution of communities [3], [9], [13], [16], [18], [25], [28] mainly by obtaining snapshots of communities at different times and identifying the changes.

Other authors, such as Chavalarias and Cointet [7] and Cointet and Chavalarias [8], proposed the application of dynamic reconstruction techniques to obtain communities' evolution over time, in order to classify terms in scientific articles, identifying patterns that allow the generation of conceptual maps based on the similarity of the analysis data, as well as on the comparison over time of communities that were born, merged with others, divided themselves or died, to see their evolution and reconstruct dynamic conceptual maps from their structure to represent a historical visualization of knowledge from the data and recognize trends in the life cycle of the communities.

III. MATERIALS AND METHODS

A. METHODOLOGY

The Cross Industry Standard Process for Data Mining (CRISP-DM) methodology [6] had been used to guide the detection of communities as a data mining problem. The methodology includes descriptions of the normal phases of a project, the tasks required in each phase, and an explanation of the relationships between the tasks. The six phases mandated by CRISP-DM work cyclically: understanding the business, understanding the data, preparing the data, modeling, evaluation and distribution.

In this work, we use CRISP-DM methodology in the following manner:

- Business Understanding: Understand the relationship between suppliers and tenders on "Mercado Público".
- Data Comprehension: the data of "Mercado Público" that will be used in this research. Analysis must be done from the perspective of the operations and rules that govern "Mercado Público" in order to become familiar with the data. Afterwards, the data must be visualized to understand the context of the problem and validate its quality in view of the project objectives.
- Data Preparation: the necessary attributes of tenders, suppliers and their relationships must be identified. Filters must be applied to the tenders with the aim of selecting representative data of the proposed problem. If there is null or incorrect data, it must be properly prepared.

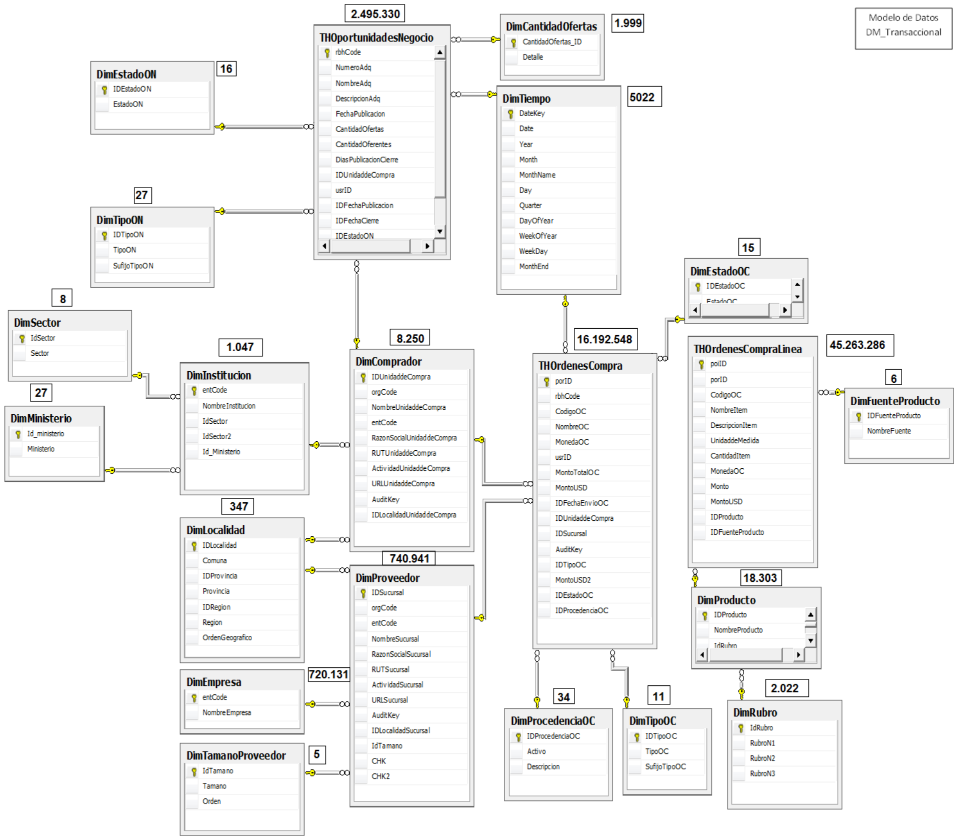


FIGURE 1. Data model of the data-mart of "Mercado Público". Source: Transparency Portal of Chile.

- Modeling and Evaluation: a model must be selected for the data using proper representation.

B. TOOLS

In this study, we used the following tools to help with the different stages of the Methodology: Microsoft SQL Server® 2008 R2 SP2, the Tableau Software Desktop 64Bit 9.2.6 https://www.tableau.com/ and the Gephi Software http://gephi.org/ 8.0.2, a free and open source visualization and exploration software for all kinds of graphs and networks.

C. DATA SET DESCRIPTION

To characterize the public procurement market we use a data mart (see Fig. 1) with the transactional data from 2009 to 2016 obtained from "Chile Compra" (http://portaltransparencia.cl/. We only work with tenders of larger size (type B2, H2, I2, LP, LQ, LR) awarded to more than one supplier.

D. METHOD

We represent the list of suppliers and awarded tenders as an undirected bipartite graph where the weights of the links correspond to the percentage of the tender awarded. In order to detect communities, the bipartite graph is projected onto a monopartite graph of suppliers where a modularity algorithm [4] must be applied and a visualization software should be used. The exercise is repeated in different years and the results of the graphs obtained are compared to each other in order to evaluate the variation and evolution of the characteristics of the communities over time.

To do this, in this study we extend the model proposed by [29] to include the relationship between suppliers that participate together in public procurement tenders. Then, the graph represents the relationship between suppliers in order to find collaborative communities of those who participate together in the same tender, considering attributes such as the item of the products involved in the transaction.

IV. RESULTS

A. BUSINESS UNDERSTANDING

Table 3 shows the number of tenders, buy orders, besides of the organisms and suppliers which were present in the Chilean Public Market for the seven years period under study.

Table 3 shows the number of tenders, buy orders, organizations and suppliers which were present in "Mercado Público" for the seven year period under study.

Table 4 shows the number of tenders greater than 1,000 UTM (i.e.: B2, H2, I2, LP, LQ and LR type) that were awarded between January 2009 and October 2016, when a

TABLE 3. Business understanding.

# Records	Description
2.495.330	Tenders
16.192.548	Buy orders
8.250	Organism
740.941	Suppliers
5	Suppliers' size
2.022	Product type

total of 86,815 transactions were made between suppliers and buyers.

TABLE 4. Amount of tenders per year and month.

Month	2009	2010	2011	2012	2013	2014	2015	2016	Total
Jan.	196	532	650	785	800	968	1144	1057	6132
Feb.	577	539	566	764	747	869	1086	1266	6414
March	602	465	659	754	725	929	1414	1177	6725
April	627	402	620	782	914	931	1233	993	6502
May	570	438	641	862	916	842	1232	1376	6877
June	682	420	708	875	830	945	1318	1241	7019
July	729	376	720	1042	1094	971	1358	1040	7330
Aug.	730	506	871	1068	1041	920	1172	1192	7500
Sept.	768	609	912	1133	1003	1025	1169	753	7372
Oct.	673	703	937	1147	1100	1265	1226	233	7284
Nov.	711	930	1621	1118	1027	1561	1524	NA	8494
Dec.	892	1080	1314	1156	1270	1758	1696	NA	9166
Total	7757	7000	10219	11486	11467	12984	15572	10330	86815

For this study we analyze tenders by year. Since we do not have all the data for 2016, we selected 2015 as the end year. To evaluate evolution over time, we will compare the snapshot of 2015 against the snapshots of the years 2014, 2013 and 2012.

B. DATA UNDERSTANDING

Using Tableau Software, we performed the exploration and validation of the quality of the data by reviewing the links between the tables of the relational model (see Fig. 1), verifying the existence of reference data, obtaining information necessary to understand the context of the study.

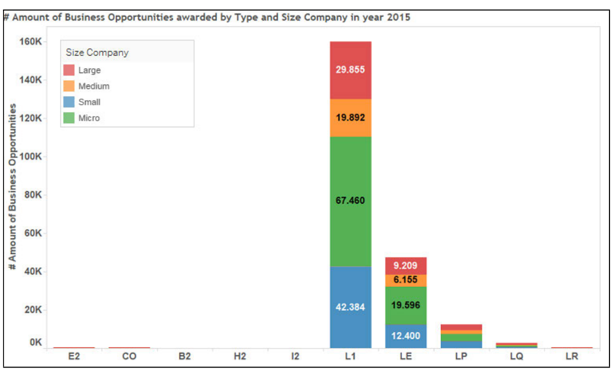


FIGURE 2. Number of tenders awarded by type in year 2015.

Fig. 2 shows that in 2015 the L1 tender type was the most awarded compared with others tenders types classified as large tenders. It important to note that the micro enterprise type was the most active.

In Fig. 3 we show the number of awarded tenders for all the years under study, separated according to company size, where distribution is similar between the large, small and micro enterprises. Note that the medium-sized companies have a slightly smaller share in distribution. In Fig. 4, we show tenders according to purchase order amount awarded according to the size of the company and year.

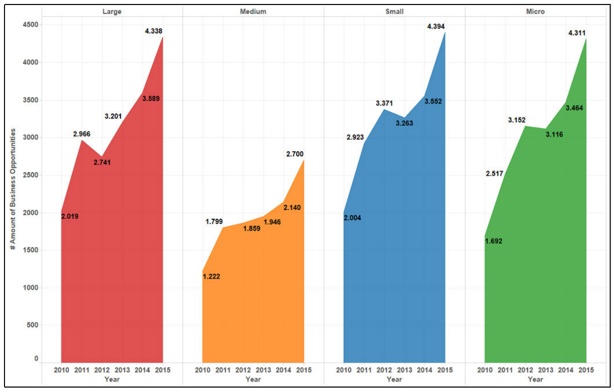


FIGURE 3. Number of awarded tenders whose purchase amount was greater than 1,000 UTM, separated by year and according to the size of the companies.

In Fig. 3 we can see that the number of tenders over 1,000 UTM is distributed evenly across companies of different sizes. However, in Fig. 4 the distribution by amount is shown. Thus, we can see that larger percentage of tenders (in terms of the amount of money awarded) are granted to large companies.

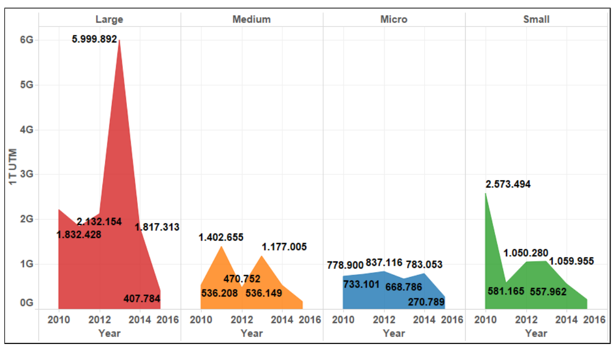


FIGURE 4. Sum of amount in UTM of tenders awarded with purchase amount greater than 1,000 UTM, according to the size of the companies and the closing date of each tender.

At this stage, we concluded that the data obtained was sufficient to follow the CRISP-DM methodology in performing a study of the Chilean public procurement market.

C. DATA PREPARATION

The graph is based on the structure of nodes described in Table 5 and the structure of the links defined in Table 6, and considers the attributes necessary to represent the graph model for the detection of communities, their characterization and their evolution over time.

TABLE 5. Node structure.

Attribute	Description
ID	Node identifier
Label	Node label [Tender/Company/Supplier Name]
Size	Company size
Type	Node type [Tender/Supplier]

TABLE 6. Bipartite link structure.

Attribute	Description
Source	ID Tender
Target	ID Supplier
Weight	Link Weight: Value Purchase order
Year	Tender year
Sector	Sector of the Buyer organism of the tender
Item	Products type in Purchase order
Type	Indirect [not directed graph]

The extraction of the data from the data mart was performed using an SQL Script, and filters were applied to obtain only the largest tenders, i.e., those of types B2, H2, I2, LP, LQ, and LR, with a specific status (''awarded''), whose orders were accepted by the purchasing agency, where the tender was awarded to more than one supplier and the award was in 2012, 2013, 2014 or 2015. We did not consider tenders awarded to only one supplier because we are trying to identify those tenders that were awarded to a group. Anomalous cases were cleaned up (for instance, the tender with ID number 6936300, since it was an outlier with almost 10,000 purchase orders to different suppliers). It is important to consider the quality of the key data, such as the size of each supplier company, the sector, and the subject of the purchase orders when examining the result of the analysis. For the case study, the definition and categorization of the data are given by the data mart, therefore, the validity of the results depends directly on the quality of the data given by the Transparency Portal. Special characters that might cause problems when reading the information, such as backslashes, semicolons, tab characters, and commas, were excluded.

D. MODELING

We model the problem as a Bipartite Graph $G(\nu, \xi)$, where the vertices $\nu = L \cup P$ of the graph consist of two disjoint sets $L = (L_1, L_2, L_3, \ldots L_m)$ and $P = (P_1, P_2, P_3, \ldots P_n)$ which represent tenders and suppliers, respectively. The set of edges ξ consists of tuples (L_j, P_k) that link together the supplier P_k with its awarded tender L_j. Moreover, $W(L_j, P_k)$ is the purchase order amount awarded P_k in the tender L_j. The resulting nodes and links are shown in Table 7.

TABLE 7. Data extraction from the dataset.

Tender Nodes type	12.222
Supplier Nodes type	7.406
Nodes total	19.628
Link Supplier-Tender	598.408

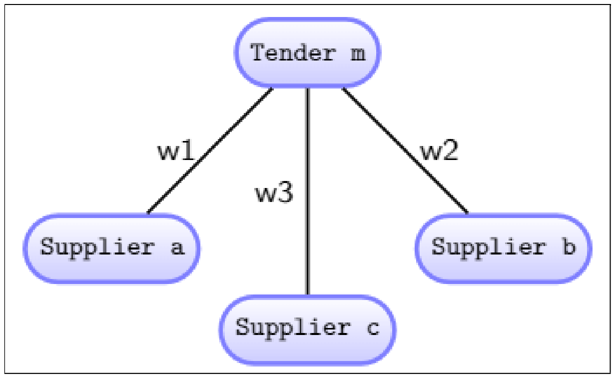


FIGURE 5. Bipartite Graph Representation: Bidding - Supplier.

The relation between suppliers and tenders is represented as a bipartite graph (see Fig. 5), where tenders and suppliers are represented as nodes, suppliers that are awarded tenders are represented as links between tenders and suppliers, and the purchase order amount awarded to each supplier is represented as the weight assigned to that link.

Algorithm 1 describes the procedure to project the bipartite graph onto a monopartite graph, eliminating bidding-type nodes in order to connect suppliers that are granted the same tenders. The logic of the projection is based on the definition of the link between two suppliers given by their joint participation in a tender, and the new weight of the link is determined by the average of the sum of the purchase orders amounts among the related suppliers.

Algorithm 1 Graph Projection

```
procedure SuppliersGraphProjection
    for all Tenders t do
        for all Suppliers s_j and s_k in Tenders t do
            w_j ← weight of Supplier s_j
            w_k ← weight of Supplier s_k
            w ← (w_j + w_k)/2
            if link between s_j and s_k exists then
                Update link (s_j, s_k, w)
            else
                Draw link (s_j, s_k, w)
```

Fig. 6 shows the final, monopartite graph once the graph's projection algorithm was applied.

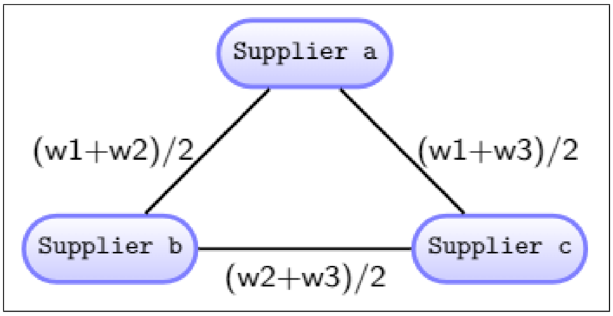


FIGURE 6. Monopartite Graph Representation: Bidding - Supplier.

TABLE 8. Data extraction considering a Monopartite graph.

Results	#
Supplier nodes type	7.406
2012 - link supplier-supplier	14.411
2013 - link supplier-supplier	16.266
2014 - link supplier-supplier	15.336
2015 - link supplier-supplier	15.338

Table 8 summarizes the results from the dataset.

We use the Gephi analysis tool to graphically represent the network of suppliers with the data of nodes and links described in the preparation stage of the data. After that, in order to better identify the communities, the modularity algorithm proposed in [4] is applied. In Fig. 7a one can observe the presentation of the graph before applying the modularity algorithm. Then, in Fig. 7b the graph is shown with the modularity algorithm applied, prior to the application of Force Atlas 2 that improves the layout. Finally, in Fig. 7c the final result is obtained after the application of both algorithms.

E. EVALUATION

The evaluation section consists of assessing how well this method and methodology allowed us to identify suppliers' communities in the different years of the study.

1) COMMUNITIES ANALYSIS

In order to characterize each community and obtain common properties [10] we identify attributes in nodes and links that allow us to describe each one of them. In the first place, we identify the number of component suppliers, the participation percentage with respect to the complete graph, and the minimum, maximum and average degree of the community. We select the item with the highest occurrence, the value obtained from the links from the products involved in the purchase order of the supplier in the tender, and divide the graph by item according to Fig. 8a. We characterize the size of the company by identifying the amount and percentage of participation in the community. We identify the central node using the graph theory measure of centrality called Eigenvector, which conditions the importance of a node within a graph according to the relevance of all the nodes that are connected to it. Finally we describe the three sectors, corresponding to the buyer organisms. Fig. 8b shows the most relevant business sector and the percentage of participation in the community.

2) COMMUNITIES EVOLUTION

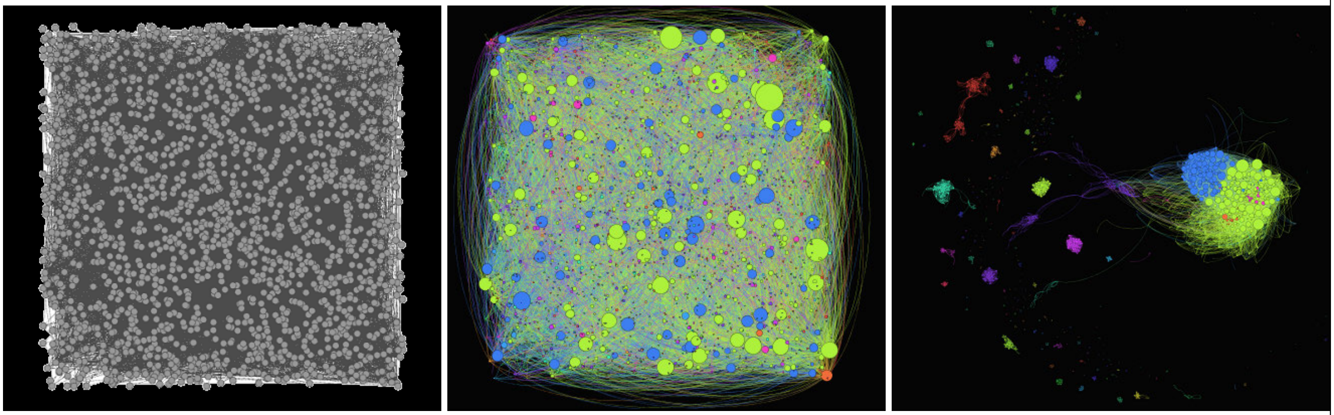
In order to analyze the evolution of the communities, we need to compare the metrics of each graph by year. Then, we must to compare the results of the five largest communities in each year and to conduct a visual inspection of the attributes and network topology of the communities obtained.

In Table 9 the results of the metrics applied to the network obtained after the application of the algorithms to the supplier graphs for the years 2012, 2013, 2014 and 2015 are shown. Over time, we can see that the number of participants increased 4% between 2012 and 2013, 3% between 2013 and 2014 and 5% between 2015 and 2014. The density indicates that the graph, in general, has few connections and the diameter (the maximum distance between two nodes) shows generally low connection between suppliers. Regarding communities: 279, 280, 331 and 352 supplier communities were detected in the years 2012, 2013, 2014 and 2015, respectively. The community structure with the lowest value occurs in 2013 with a modularity value of 0.361, lower by 0.286 if we calculate the average of the difference with respect to 2012, 2014 and 2015. This may be the cause of the increase in the number of links and the value of the average grade detected in the year 2013. This could be caused by an increase in the amount of joint participation from suppliers, but not by different suppliers.

TABLE 9. Evolution metrics of Suppliers' Communities in the Public Market.

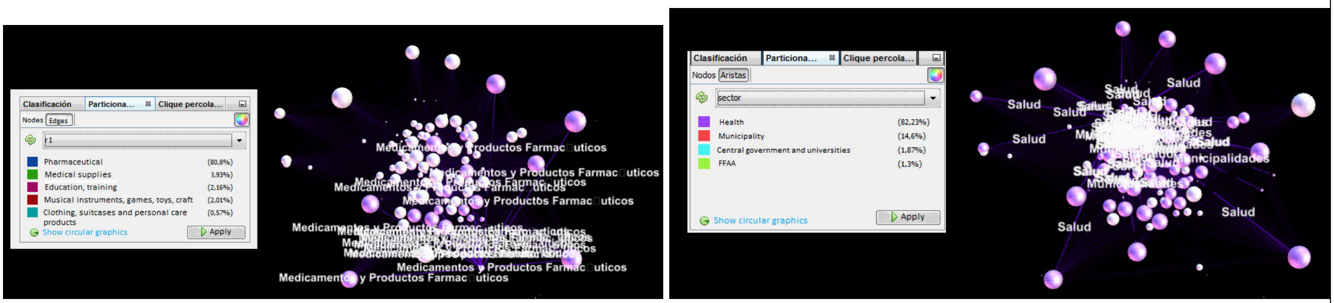
Metrics	2012	2013	2014	2015
# Tender Nodes	1.830	1.903	1.967	2.074
# Links	13.832	15.559	15.336	14.734
# Communities	279	280	331	352
Modularity	0,660	0,361	0,577	0,704
Network diameter	11	14	10	12
average grade	15,1	16,352	14,981	14,8
Density	0,008	0,009	0,008	0,007
Large of average path	3,138	3,565	2,61	3,267
average coefficient of clustering	0,866	0,866	0,863	0,859
centrality of eigen vector	0,052	0,030	0,055	0,031
Micro	51%	50%	52%	54%
Small	24%	26%	26%	25%
Medium	13%	12%	12%	11%
Large	12%	12%	11%	10%

In Fig. 9, we apply a filter of node degree (greater than 100). We can visualize that the nodes with 100 or more



(a) Graph in Phase 0 (b) Graph in Phase 1 (c) Graph in Phase 2

FIGURE 7. **Visualization of Graph at year 2015. Phase 0: Graph before of applying the modularity algorithm and Force Atlas 2. Phase 1: Graph with modularity algorithm applied before Force Atlas2. Phase 2: Graph with modularity and Force Atlas 2.**



(a) Items type Graph 2015 (b) Visible Business Sectors in the Graph 2015

FIGURE 8. **Characterization of the Graph community at year 2015 showing the most relevant item and business type respectively.**

links in the suppliers in the year 2013 belong to the largest community. This means that the joint participation of these suppliers occurs most frequently.

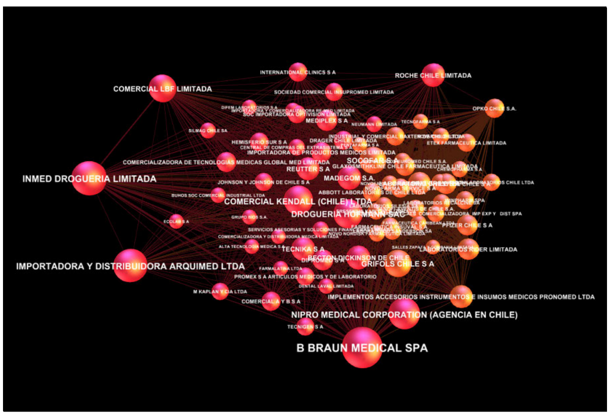


FIGURE 9. **Visualization of suppliers nodes during year 2013, whose grade is greater than 100.**

The greatest modularity value occurred in the year 2015, reflecting solid connections between suppliers within each community, but scarce connections between suppliers from the other communities. If we compare the distribution of company sizes in the 2012-2013 period with the 2014-2015 period, we can see an increase in the participation of micro companies and a decrease in large and medium-sized

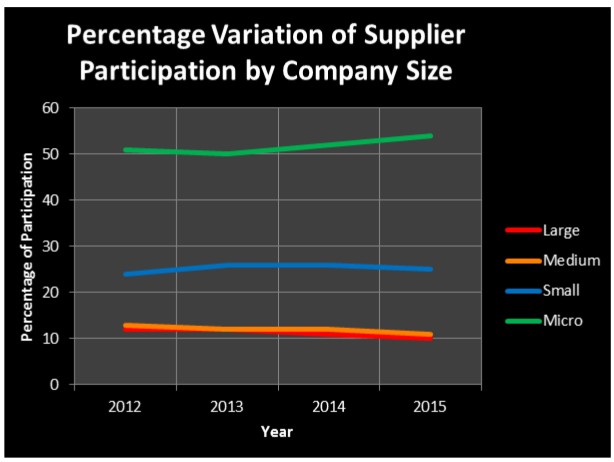


FIGURE 10. **Participation comparison of suppliers by company size in communities during years 2012, 2013, 2014 and 2015.**

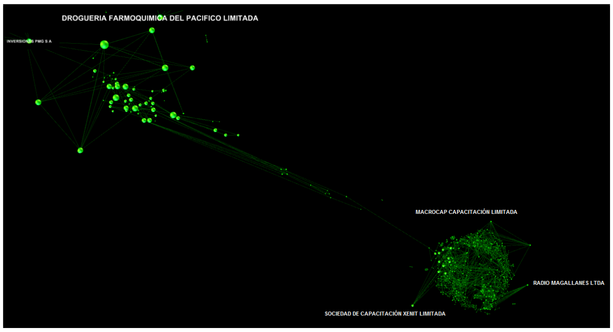


FIGURE 11. Visualization of suppliers nodes of year 2013, whose grade is less than 100 and the size of the providers corresponds to the micro companies.

companies' participation in the second period, according to Fig. 10. Moreover, in Fig. 10 we can see greater participation of the small and micro companies compared to the medium and large companies, which may indicate that the suppliers that participate in the community are more likely to win a bid if they participate together; however, they co-exist in smaller communities. For example, the average participation of micro enterprises in the communities is 52% for the four years analyzed, but these companies participate mainly in smaller communities with a lower node grade, which means that there are several markets that are niches, as we can see in Fig. 11, where suppliers such as: "Radio Magallanes Ltda", "Sociedad de Capacitación Xenit Limitada" and "Macrocap Capacitación Limitada" belong to the group of smaller communities with fewer connections.

Fig. 12a, Fig. 13a, Fig. 14a and Fig. 15a show the supplier communities through the years 2012 to 2015, indicating distribution by community as well as distribution by company size (see Fig. 12b, Fig. 13b, Fig. 14b and Fig.15b).

3) DISCUSSION

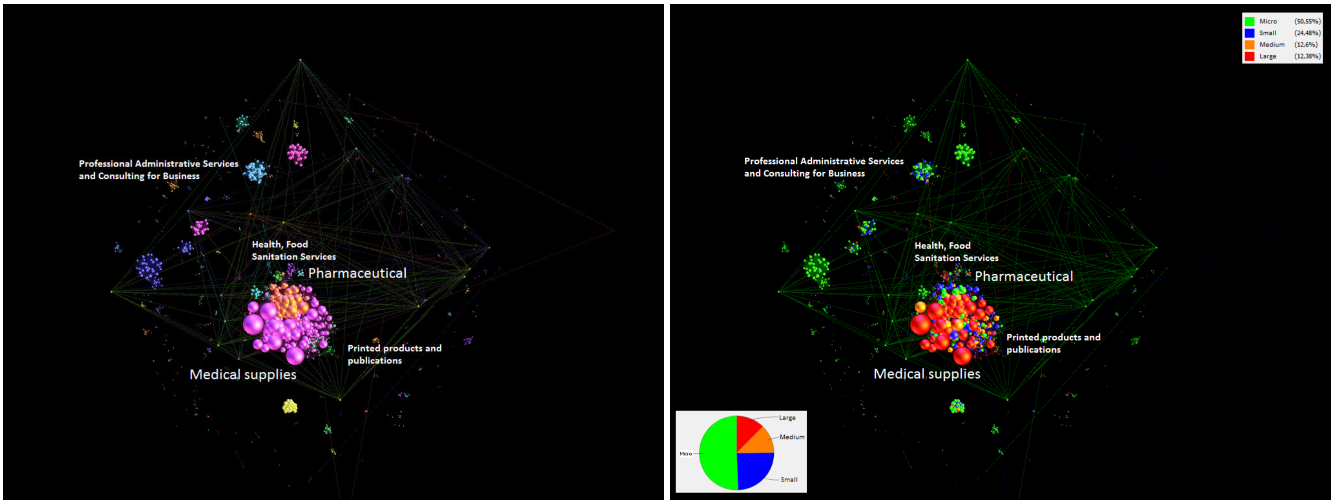
We obtain the following information from the five most relevant communities for each year:

- The supplier size with the greatest participation within the five communities analyzed between the years 2012 and 2015 was the micro company category, with an average of 52%, followed by the small business category with an average participation of 25%, and the medium-sized companies with 12%,
- The lowest community participation was obtained by large company category, with an average participation of 11% according to data obtained from Table 9.
- The temporal trend showed that the micro size companies have increased their participation, unlike the medium and large size companies, which have shown a decrease.
- The community providing Equipment, Accessories and Medical Supplies has remained the most relevant in the four years under analysis, with an average participation of small and micro businesses between 2012 and 2015 of 36% and 29%, respectively. Micro companies showed a growth trend over time from 27% to 32%.
- The community related to Medications and Pharmaceutical Products was the second-most relevant in the years 2012, 2014 and 2015; even though it fell to fifth place in 2013. It is the only community where the largest company occupies the highest average participation (40% of total suppliers between 2012-2015).
- The community related to Clothes, Suitcases and Personal Cleaning Products was not relevant in 2012, but went up in 2013 to the fourth place of importance measured by community size, maintaining itself at the fifth place in 2014 and 2015 with the greater participation of the Armed Forces sector: close to 70% on average for those three years, entering 2015 with that sector's participation percentage in the community at 95%.
- The community related to Consultancy for Business Management and Training saw a contraction if we compare 2012 and 2013 to 2014 and 2015, showing a decrease on average of suppliers in the community from 91 to 82. In the years 2012, 2013 and 2014 the largest participating sector corresponded to Central Government-Universities, but in 2015 the sector of greatest relevance to this community corresponded to the Health sector.
- The community related to Health, Food Sanitary Services appears in the last place within the five largest communities in 2012, and in third place in 2013, with a greater share of the Armed Forces sector in 2012 and the Health sector in 2013, with an increase in the number of suppliers from 52 to 69, but disappearing the following years.
- The community related to **Forestry Fisheries and Wildlife-related Agricultural Services** was born in 2014 with 97 suppliers and disappears in 2015,
- The community related to Musical Instruments, Games, Toys, Crafts, and Education Materials, Accessories and Supplies was also born in 2015 with 65 suppliers.
- The community related to Print Products and Publications was born in 2012 with 65 suppliers, disappearing the following years.

In Fig. 16 we can see the evolution of participation percentage by sector for the Medical Equipment, Accessories and Supplies community, observing that the Health sector had greatest participation for all four years analyzed, followed by the Municipal Government sector, both with a slight tendency toward decreasing participation.

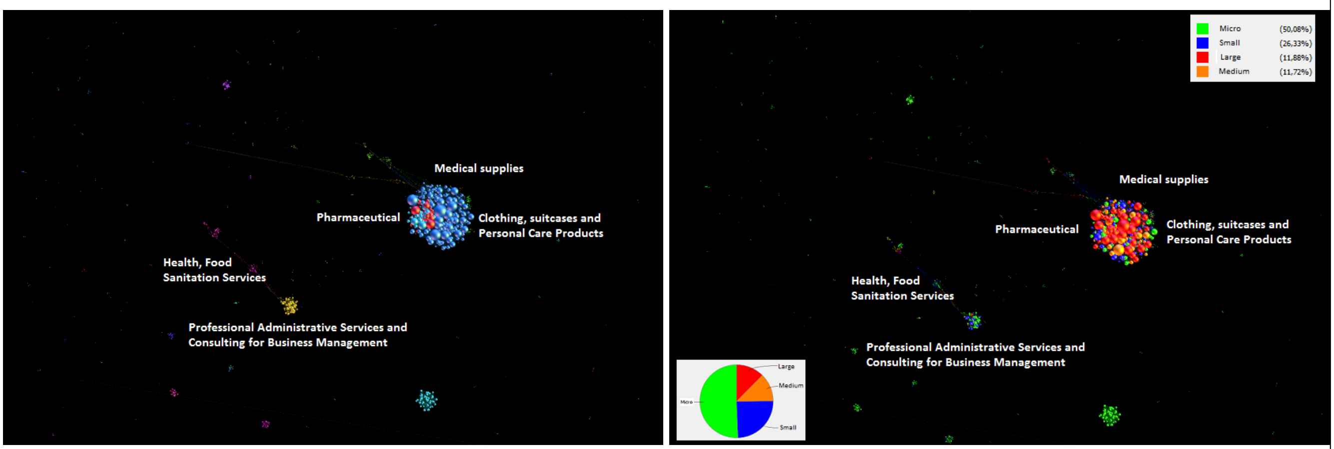
In Fig. 17 we can see the evolution of participation by company size, where the Medical Equipment, Accessories and Supplies community has the greatest participation with a positive trend, and medium and large size companies have a similarly low participation with a negative participation trend.

By comparing our technique and results with the closest related works, we can conclude that, differently from [29], we were able to detect suppliers' communities and make a temporal analysis of them. While other works have previously used this technique to detect communities in a different area



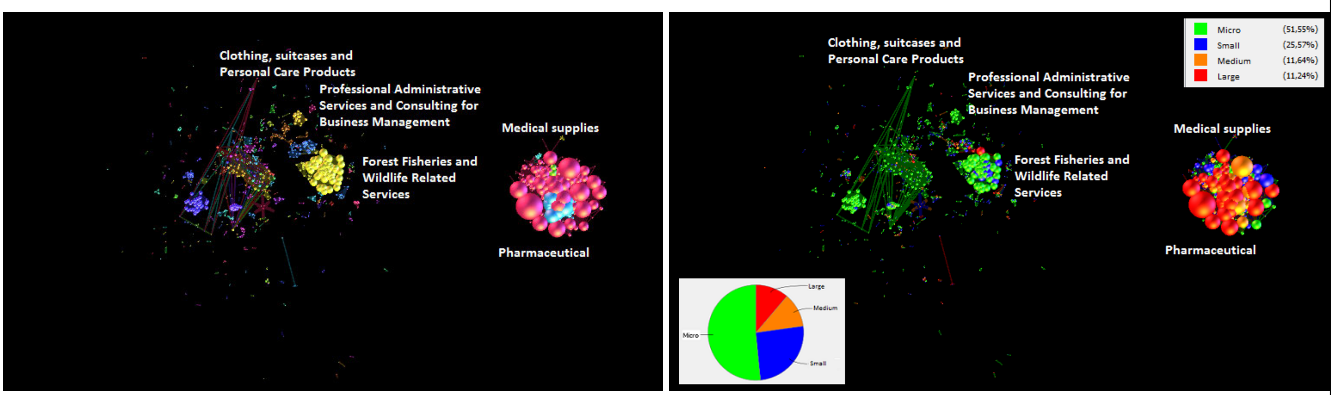
(a) Year 2012 - Distribution by Communities
(b) Year 2012 -Distribution by Company Size

FIGURE 12. Suppliers' communities during year 2012, colored by modularity and by company size, highlight of the five most relevant.



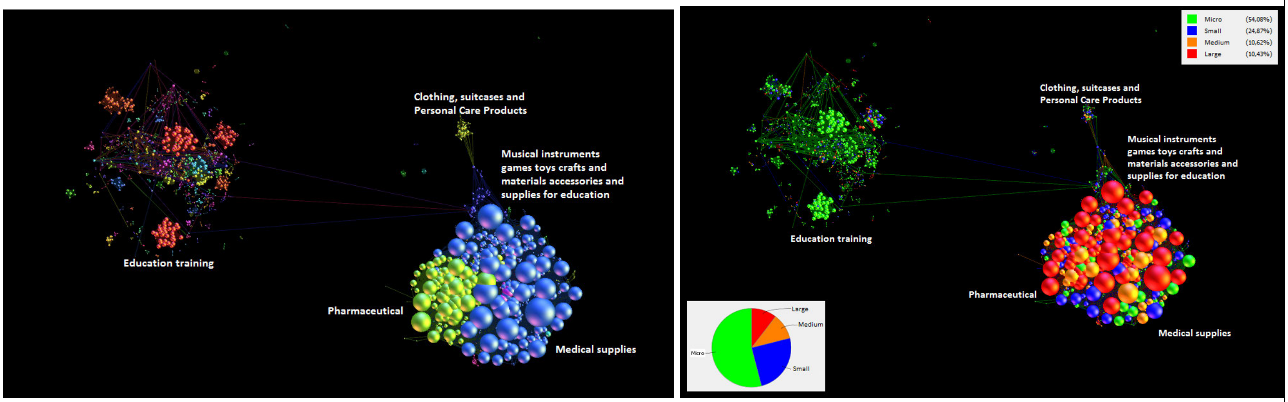
(a) Year 2013 - Distribution by Communities
(b) Year 2013 - Distribution by Company Size

FIGURE 13. Suppliers' communities during year 2013, colored by modularity and by company size, highlight of the five most relevant.



(a) Year 2014 - Distribution by Communities
(b) Year 2014 -Distribution by Company Size

FIGURE 14. Suppliers' communities during year 2014, colored by modularity and by company size, highlight of the five most relevant.



(a) Year 2015 - Distribution by Communities (b) Year 2015 -Distribution by Company Size

FIGURE 15. **Suppliers' communities during year 2015, colored by modularity and by company size, highlight of the five most relevant.**

[15], in our work we were able to perform a temporal analysis of these communities.

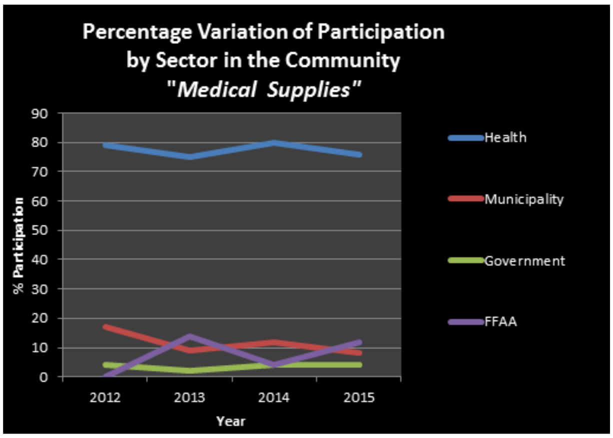


FIGURE 16. **Variation in the % of participation by sector in the Equipment, Accessories and Medical supplies, between the years 2012, 2013,2014 and 2015.**

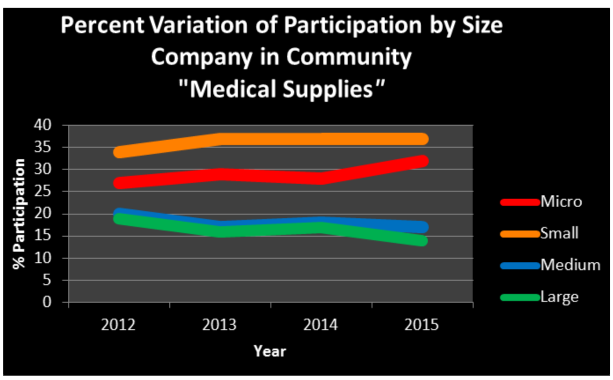


FIGURE 17. **Variation in the % of participation of companies of different size in the Equipment, Accessories and Medical Supplies' community.**

V. CONCLUSION AND FUTURE WORKS

As a result of the work, supplier communities were detected in the period under study, so we can conclude that it is possible to reconstruct the public procurement market. To perform the detection, we obtained data from ''Mercado Público'', which were analyzed and transformed to a bipartite graph that represented the relationship between the suppliers that were awarded the same tender. These communities were characterized according to the size of the involved companies, the sector of the buyer agency and the products associated with the purchase orders. We made a description based on the structure of the resulting graphs of the five largest communities for the period 2012–2015, where we observed a greater participation in the number of suppliers of small and micro enterprises compared to medium and large size companies. Regarding the presence in the community of medium and large size companies, although they have a smaller share of the number of suppliers in the detected communities, the nodes that represent them have a greater degree size, which suggests that, as opposed to small and medium size companies, they have a greater market share, that is, greater interaction with other suppliers in particular segments. Small and micro companies participate in smaller communities, unlike large and medium companies. The community of ''Medicines and Pharmaceutical Products'' that has remained within the five largest communities detected, is the only community in which large companies show the largest average participation between 2012-2015, with 40% of total suppliers. On the other hand, the most relevant community corresponded to ''Medical Equipment, Accessories and Supplies'' with an average participation of the small and micro company between 2012-2015 of 36% and 29% respectively, with the micro company category's growth trend over time from 27% to 32%. These observations indicate that when analyzing each community, it is possible to identify participation trends by sector, which can give an indicator of which items are more accessible for small and micro enterprises.

Regarding the temporal analysis of the communities, we conclude that the temporal trend showed that the micro companies have increased their participation, while the medium and large companies have seen their participation decrease. In general, we observe that when descriptive metrics of the communities in each year are compared, it is possible to identify trends in the evolution of the behavior of the suppliers with respect to their community structure, the strengthening of communities over time, the birth and death of other communities, and the variation of participation regarding the size of the suppliers. By changing the parameters of the model, different aspects of the supplier communities may be observed (for example, the suppliers' behavior in a specific sector). For a more detailed analysis, we propose to filter the parameters (year, sector, company size, and item) to obtain information about the communities and analyze results and the evolution of each community over time.

In a future work, we plan to make a deeper analysis by studying each specific community and its evolution for all available types of tenders. It would also be interesting to directly analyze the bipartite network in order to detect communities by considering the weight of the link between the supplier and the bidding given by the amount of related purchase orders lost when projecting the network to a monopartite graph of supplier type nodes. In that case, the network will be composed not only of suppliers but incorporate the tenders, which will allow us to obtain other metrics of interest from the relationship between suppliers and tenders awarded, such as the amount awarded to each community, valuing their participation and evolution. Regarding the structure of the graph, we suggest to investigate the application of algorithms that allow the detection of communities in overlapping networks, identifying nodes that may belong to different communities, in order to detect suppliers that could be intermediaries between different communities. Finally, we also propose in future to apply the technique of dynamic reconstruction to obtain the evolution over time of the suppliers' communities, information that can be useful to classify suppliers by category and allow improved searches of historical data, identifying patterns that permit the generation of conceptual maps based on the similarity of analysis data. In addition, we expect to extend this method to obtain specific indicators in order to make a comparison over time of the communities that are born, merging with others, dividing, or dying.

REFERENCES

[1] M. A. Aljemabi and Z. Wang, "Empirical study on the evolution of developer social networks," *IEEE Access*, vol. 6, pp. 51049–51060, 2018.

[2] A. Fernandes, P. C. T. Gonçalves, P. Campos, and C. Delgado, "Centrality and community detection: A co-marketing multilayer network," *J. Bus. Ind. Marketing*, vol. 34, pp. 1749–1762, Feb. 2020.

[3] S. Asur, S. Parthasarathy, and D. Ucar, "An event-based framework for characterizing the evolutionary behavior of interaction graphs," *ACM Trans. Knowl. Discovery Data*, vol. 3, no. 4, pp. 16:1–16:36, Nov. 2009.

[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, Oct. 2008, Art. no. P10008.

[5] I. Y. Blount and J. A. Hill, "Supplier diversification by executive order: Examining the effect reporting compliance, education and training, outreach, and proximity to leadership have on government procurement behavior with minority business enterprises," *J. Purchasing Supply Manage.*, vol. 21, no. 4, pp. 251–258, Dec. 2015.

[6] C. Shearer, "The CRISP-DM model: The new blueprint for data mining," *J. Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

[7] D. Chavalarias and J.-P. Cointet, "Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study," *Scientometrics*, vol. 75, no. 1, p. 20, 2008.

[8] J. P. Cointet and D. Chavalarias, "Multi-level science mapping with asymmetrical paradigmatic proximity," *Netw. Heterogeneous Media*, vol. 3, no. 2, pp. 267–276, 2008.

[9] D. J. Fenn, M. A. Porter, M. McDonald, S. Williams, N. F. Johnson, and N. S. Jones, "Dynamic communities in multichannel data: An application to the foreign exchange market during the 2007–2008 credit crisis," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 19, no. 3, Sep. 2009, Art. no. 033119.

[10] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, Feb. 2010.

[11] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[12] P. M. Gleiser and L. Danon, "Community structure in jazz," *Adv. Complex Syst.*, vol. 6, no. 4, pp. 565–573, Dec. 2003.

[13] J. Hopcroft, O. Khan, B. Kulis, and B. Selman, "Tracking evolving communities in large linked networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 1, pp. 5249–5253, May 2004.

[14] I. Lee and B.-W. On, "Community detection for large scale social network sites," in *Proc. IEEE Southeastcon*, Mar. 2012, pp. 1–5.

[15] R. K.-X. Jin, C. D. Parkes, and J. P. Wolfe, "Analysis of bidding networks in ebay: Aggregate preference identification through community detection," in *Proc. AAAI Workshop*, Jan. 2007, pp. 1–8.

[16] M.-S. Kim and J. Han, "A particle-and-density based evolutionary clustering method for dynamic networks," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 622–633, Aug. 2009.

[17] B. Krishnamurthy and J. Wang, "On network-aware clustering of Web clients," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 30, no. 4, pp. 97–110, Oct. 2000.

[18] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and L. B. Tseng, "Facetnet: A framework for analyzing communities and their evolutions in dynamic networks," in *Proc. 17th Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2008, pp. 685–694.

[19] F. Liu and G. Xie, "A fast algorithm for community detection of network systems in smart city," *IEEE Access*, vol. 7, pp. 51856–51865, 2019.

[20] K. Loader and S. Norton, "SME access to public procurement: An analysis of the experiences of SMEs supplying the publicly funded UK heritage sector," *J. Purchasing Supply Manage.*, vol. 21, no. 4, pp. 241–250, Dec. 2015.

[21] J. Moody and D. R. White, "Structural cohesion and embeddedness: A hierarchical concept of social groups," *Amer. Sociol. Rev.*, vol. 68, no. 1, pp. 103–127, 2003.

[22] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 2, pp. 404–409, Jan. 2001.

[23] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, Jun. 2006.

[24] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, Feb. 2004, Art. no. 026113.

[25] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, Apr. 2007.

[26] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005.

[27] P. K. Reddy, M. Kitsuregawa, P. Sreekanth, and S. S. Rao, "A graph based approach to extract a neighborhood customer community for collaborative filtering," in *Proc. 2nd Int. Workshop Databases Netw. Inf. Syst. (DNIS)*. Berlin, Germany: Springer-Verlag, 2002, pp. 188–200.

[28] J. Sun, C. Faloutsos, C. Faloutsos, S. Papadimitriou, and S. P. Yu, "Graphscope: Parameter-free mining of large time-evolving graphs," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2007, pp. 687–696.

[29] N. M. Vafopoulos, I. Anagnostopoulos, M. Meimaris, M. Klonaras, I. Xidias, A. Papantoniou, G. Vafeiadis, V. Loumos, and G. Alexiou. (Mar. 2013). *Network Analysis of Public Expenditure: The Case of Greece*. [Online]. Available: https://ssrn.com/abstract=2229303, doi: 10.2139/ssrn.2229303.

[30] P. Zhang, J. Wang, X. Li, M. Li, Z. Di, and Y. Fan, "Clustering coefficient and community structure of bipartite networks," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 27, pp. 6869–6875, Dec. 2008.

FABIOLA HERRERA LEIVA received the B.Eng. degree from the Pontificia Universidad Católica de Valparaíso, Chile, and the master degree in informatics engineering from Andres Bello University, Viña del Mar, Chile, in 2018.

Since 2004, she has been in software development with the Automotive Industry, as a Senior Full Stack Software Engineer. She is currently a Project Manager with various development projects. She also trains in data-science, specifically in tools, such as SQL, Phyton and R, and applied to graph analysis. Her research interests include analysis of public data, especially open government data and data analysis.

ROMINA TORRES (Member, IEEE) received the B.S. and M.Sc. degrees in informatics engineering and the D.Eng. degree in informatics from Federico Santa María Technical University, Valparaíso, Chile, in 2001, 2003, and 2014, respectively. From 2001 to 2008, she worked with major international companies, such as Motorola, and Software AG. Since 2013, she has been a Professor with the Engineering Faculty, Andres Bello University, Viña del Mar, Chile, where she is currently the Director of the Computer Science Program. Her research interests include intelligent systems and their applications to support decisions in several areas.

ORIETTA NICOLIS received the degree in economics (mathematics and statistics) from the University of Verona, Italy, in 1995, and the Ph.D. degree in statistics from the University of Padua, Italy, in 1999. She held a Postdoctoral Fellowship position in statistics with the University of Brescia, Italy, for a period of two years. From 2002 to 2012, she was a Researcher with the Department of Information Technology and Mathematical Methods, University of Bergamo, Italy. From 2012 to 2018, she was a Professor with the Universidad de Valparaíso. Since August 2018, she has been a Professor with the Engineering Faculty, Andrés Bello University, Viña del Mar, Chile, where she is currently the Director of the master's Program in computer science. She is the author of many scientific publications. Her research interests include study of time series models, spatial and spatial-temporal models applied to environmental phenomena, machine learning methods, big data, computer science, and wavelet analysis. She is a member of the international statistical societies.

RODRIGO SALAS F (Senior Member, IEEE) received the B.S. and M.Sc. degrees in informatics engineering and the D.Eng. degree in informatics from Federico Santa María Technical University (UTFSM), Chile, in 2001, 2003 and 2010, respectively. From 2002 to 2004, he was a Research Assistant with the Informatics Department, UTFSM. Since 2004, he has been with the Universidad de Valparaíso, where he is currently a Full Professor with the Biomedical Engineering School and Teaches in data mining, probability and statistics, and machine learning. He is also a Main Researcher with the Center of Research and Development in Health Engineering (CINGS-UV). He participates in several programs with the Universidad de Valparaíso, such as the Biomedical engineering, the M.Sc. degree in biomedical engineering, the M.Sc. degree in statistics, the Ph.D. degree in Statistics, and the Ph.D. degree in science and engineering for health. His research interests include computational intelligence, data science, decision support systems, intelligent systems and their applications to finance, air pollution, healthcare, and medicine.