Using Algorithms K-means and Centrality of Intermediation for Analysis of a Social Network Focused in Commercial Disclosures

Carolina Palma Pimenta Furlan
Curso de Ciência da Computação
Centro Universitário UnirG
Gurupi, Brasil

Lucas Santa Bárbara Gomes
Curso de Ciência da Computação
Centro Universitário UnirG
Gurupi, Brasil
Email: lucas.sbgomes [AT] gmail.com

Marcelo Lisboa Rocha
Curso de Ciência da Computação
Universidade Federal do Tocantins
Palmas, Brasil

David Nadler Prata
Curso de Ciência da Computação
Universidade Federal do Tocantins
Palmas, Brasil

Anderson da Silva Soares
Curso de Ciência da Computação
Universidade Federal do Goiás
Goiânia, Brasil

Clarimar José Coelho
Curso de Ciência da Computação
Pontifícia Universidade Católica de Goiás
Goiânia, Brasil

Abstract— **The advent of the Internet has enabled several means of communication between people and among them is the social networking platforms becoming a new form of relationship. Explore this new environment, it has become increasingly popular for researchers, as well as managers in general. This environment provides the formation of communities and through them it is possible to identify formation of groups through their interests and connections. The environment of this research consists of subgroup Tocantins digital social network Facebook. This subgroup that is focused exclusively on commercial disclosures, having more than 10,000 members. In this environment, it was possible to develop applications to collect information of the postings of the members of the group studied, by means of APIs that Facebook itself offers. Later in this work, we applied the algorithm k-means to the grouping of data (users with same preferences), and also to measure centrality of intermediation where they revealed the existence of States with greater influence in your posts, and products or services most viewed within the subgroup.**

Betweenness centrality; k-means; Facebook; social network; commercial disclosures.

I. INTRODUCTION

A social network is composed of a group of people, organizations or other social entities and their relations socially significant, as friends, work colleagues or exchange of information, which can be connected by computer networks [1]. For e-commerce, social networks have the purpose to meet and understand the needs of the consumer that is not only seeking to purchase, but in search of content and exchange of experiences, which assists the user in the process of purchasing decision [2]. With the amplitude of this scenario, companies that use e-commerce are looking for tools that indicate a greater propensity of this public for analysis of future business projections and marketing policies.

The social network analysis (SNA) [1] was addressed by Freeman in 1996, within a sociological background, describing the means by which information can be encoded. The SNA can be applied in various areas of knowledge as knowledge management, organizational development, research in economics and social sciences, epidemiology and security against terrorist attacks, among others. The majority of social networks are for personal purposes and socialization. However, the analysis of social network leads to mapping, measurement

and modeling of relationships and flow between people and groups [3].

The Tocantins digital subgroup is a private space composed of members with common interests in the social network Facebook online. It is an environment of postings classified where its members can post offers, through the posts and also receives them from other members. In this context, the subgroup Tocantins digital has all the typical characteristics of a general community formed by Facebook. Facebook offers to programers and developers an Application Programming Interface (API), which allows the development of applications that can be integrated into the environment Facebook without the involvement with the details of the original programs of Facebook's environment[4, 5]. This API provides resources for the collection of information from Facebook platform that are published by members of the Tocantins digital [6].

In the world wide web (Web) environments as Facebook, data are unstructured [7]. The data structure of Facebook is organized in a graph that represents the relationships between a set of entities [8, 9]. A graph is a structure that contains vertices and edges. In this work, the vertices of the graph represent the classes of goods and services, while the edges represent the flow between the classes. The activities of the creation of new applications involve the creation of a new environment organized in a new structure for the storage of data extracted from Facebook. The development of new applications from the facebook platform requires data to be structured by the developer once on Facebook data are disorganized. The data used for this study are from tables with information on the postings and profiles of users that are part of the Tocantins Digital, currently with more than 10000 members.

This work presents a software developed to identify groups of similar goods or services that are more posted, highlight which States have greater influence in posts of offer for the provision of services or products and submit what products or services are most viewed the subgroup of Facebook called Tocantins digital, Using this, the algorithm k-means [8, 10, 11, 12] and the measure of centrality of intermediation [13,14,15,16].

The identification of groups of products and services is fair using k-means clustering, which separates the objects into groups according to a measure of distance or similarity between them [10]. The task of data clustering is summarized in the use of the information found in the data and their relationships to divide the data into groups, which consist of elements of a group related among themselves, but not related to elements of other groups [17], as well as, allow for the observation, recording and analysis of user behavior and their reactions [18].

To determine the extent of influence is used the centrality of intermediation, which is a measure of the influence, due to the flow of data movement, between the vertices of a graph, whose calculation involves the amount of smaller paths, among all existing, passing through a given vertex [14]. The problem of minimum path consists in minimizing the cost of crossing a graph between two vertices. The cost is given by the sum of the weights of each edge travelled. For example, the interaction between two vertices V1 and V2, V3 may depend on other vertex located in the communication path between V1 and V2. The information to be shared between vertices V1 and V2 have, for the most part, pass through the vertex V3. Therefore, the vertex V3 is present in most of the smaller existing paths in the graph and has a high centrality, to have a considerable influence by virtue of their control over the information that passes between the other vertices [19].

The processes within a social network are considered dynamic because of this network is constantly changing through the relationships between people, groups or organizations [20]. The relevance of this study is true not only for the fact that the social network Facebook currently be an environment that promotes interaction between people, but also for being considered the most accessed social network in the world [21]. Thus, in this scenario, the digital Tocantins social network will serve as a tool to demonstrate to companies that participation in e-commerce should be integrated into existing business community.

The case study proposed in this work has as methodological basis the process Knowledge Discovery in Database (Knowledge Discovery in Databases KDD), which is the process of converting raw data into useful information, combining traditional methods of data analysis with sophisticated algorithms to process large volumes of data [8] [22]. Therefore, this work involved the steps outlined below: in the first stage were conducted literature searches on the algorithms to be applied in data mining and then the tools offered by the Tocantins Digital subgroup for the extraction of data. In the second stage, the experimental data were extracted through the Facebook API tool, which provides the digital Tocantins subgroup data tables to query data through the Facebook Query Language (FQL). From these tables, information related to this subgroup posts were extracted and then stored in a database of Structured Query Language (SQL). Next, the values used as parameters were analyzed to find groups by Simple K-means API, which is an algorithm of Knowledge Analysis at the University of Waikato enviroment (Waikato Environment for Knowledge Analysis-WEKA) [23], in which makes the use of K-means to separate the groups (clusters) of classes of goods and services, according to postings made by the members of the subgroup searched. WEKA, which is at version 3.6 (Oct, 2014), is a public domain tool that can be used in KDD, it presents a collection of data mining tools and provides unsupervised classification algorithms (clustering techniques), association rules, decision trees among others.

After the process of identification of groups, other research was presented using the measure of centrality, which uses the calculation of the shortest path between the vertices as a parameter to identify which subgroup members have greater influence in posts and which products / services are more posted. With an increasing number of people exchanging information and posting some kind of product or service offering within their own social networks formed groups was possible to see a new business environment, that sparked interest by an analysis. However, it is possible, at the end of

this analysis, obtain patterns of preferences for products and services more posted, as also detect which vertices have a greater influence on your posts.

With the result of this work, managers of marketing and e-commerce area can verify through this scenario, the behavior of members of the social network. May, these managers make use of desired manner to meet the client that aims defining an adequate advertising of certain types of products or services that can be offered within the social network researched and avoiding, at certain moments, dull advertising to those people who have no interest in those products or services.

The remainder of this article is organized as follows. The next section shows how were made the data's extractions and preparations of the facebook group Tocantins Digital. Section 3 presents the computational results obtained by the application of the techniques of simple K-means and the calculation of centrality of intermediation. Finally, in section 4 are presented the conclusions and future work.

II. SOCIAL NETWORKS

The digital Tocantins subgroup allows, through its own API tools, perform the extraction of information and allow other programs to use its functions. Digital Tocantins API has functions that are used to complement applications, using data extracted from the relationships of its members and information contained in the posts, which are stored in table format.

Initially, the data can only be extracted by a member of registration in the digital Tocantins subgroup. Then you can develop for this platform provided applications such as Graph API tool (Graph API), which allows you to see the elements of the graph and the connections of the studied subgroup. Tools APIs are open source voluntarily by manufacturers of programs to allow other developers to take ownership of interesting functions and add them to their own applications [24].

The Graph API is based on Hypertext Transfer Protocol Hypertext Transfer Protocol (HTTP) for data query and allows access to the social graph digital Tocantins, which is uniformly objects in the graph and the connections between them [6]. Thus, the posts of members is constructed a graph G consists of a set of vertices V, represented by the members, products and services. The connections between the vertices are represented by edges E [25]. Through this API, were extracted data of the first posts to those published in December of 2013.

Facebook has another API that allows access to the social graph, called Facebook Query Language (Facebook Query Language, FQL), which uses a query language (queries) similar to Structured Query Language (Structured Query Language, SQL). The FQL allows you to run multiple calls on a single query and also the possibility to choose the response return format, and Java Script Object Notation (Java Script Object Notation, JSON) or Extensible Markup Language (eXtensible Markup Language, XML).

To allow developers a set of functions to access the digital Tocantins API, including access to all features of the Graph API and FQL, Facebook has the opensource application Software Development Kit (Software Development Kit, SDK). The SDK is constantly used to perform operations as an application administrator, but it can also be used to perform operations on behalf of the current user session. By eliminating the need to manually manage access tokens, the SDK simplifies the process of authentication and authorization of users to your application. The SDK's official to integrate with Facebook is available for iOS, Android, JavaScript and PHP [6].

The APIs are not only a means to open source to enrich an application. They are also a potential source of revenue derived for the editor authorizing the use of his [24] system.

A. Tocantins Digital Data Extraction

Data were obtained through a data set extracted by the digital Tocantins subgroup API tool. This platform provides several tables for consultation purposes to data posted by its users through the FQL. In this work we use only the stream table that contains information of the posts. The stream can be found through the FQL to return a list of information of posts a user so that the user must grant access permission (read_stream and read_insights) to the app for its extraction.

On the home page of the digital platform Tocantins, developers.facebook.com, the App cfurlan was created which gets the user identification ID and a secret app which is the password generated by the subgroup own. Thus, it was possible to set the home page cfurlan.grupog3brasil.com.br, which is a particular sub-field of researchers to perform data extraction. However, it is possible to use any sub-domain to carry out extraction.

Through this access, with the permissions granted read_mailbox (view the user's inbox), read_stream (viewing posts) and read_insights (retrieves metrics for all pages and domains owned by the user), we used the PHP SDK, which have two classes, one cfurlan class that provides a concrete implementation that uses PHP sessions to store user IDs and access tokens, and the BaseFacebook class, which provides access to the platform. This class provides most of the necessary functionality, but it is an abstract class, because it is designed to be a subclass so that you can define how the data should be stored in the application. Digital Tocantins class implements these abstract methods using PHP [6] sessions. After extraction of the data, they were stored in tables in the same App site in a MySQL database management system.

B. Data Cleaning

After data extraction, starts a cleaning process of null values through a SQL query. Raw information that were collected from Tocantins Digital subgroup from stream table of FQL, were after stored on a new table of cfurlan App on MySQL database. Later, was made a SQL query that search on message field by null or equal values, turning possible delete posts identified on this query from App's database. Equal values corresponds to the insertion of the same post several times by the same member.

C. Data Reduction and Projection

In the App cfurlan database was created the following SQL tables, class_products, class_services, stream and client_product.

To assist in understanding the class_products table, the computer categories were created, vehicles, appliances, appliances, furniture, mobile phones and to sort the products offered in the classifieds of researched subgroup. As an example, there is a post that contains computer sales information which is classified in the category type of data processing.

In class_services table was just created class service, which ranks as a painter services providers, bricklayer, electrician, plumber, domestic, cleaner, driver, carpenter among others.

TABLE I. TABLE CLIENT-PRODUCT.

actor_id	p	p2	p3	p4	p5	p6
5251521XX	0	0	0	1	0	0
5334357XX	0	0	0	0	0	0
5847576XX	0	14	0	0	0	0
6011936XX	0	0	0	0	0	0
6142021XX	0	1	0	0	0	1
6295014XX	1	0	12	0	0	15
6336199XX	0	0	7	0	0	0

In the table stream were listed 6,579 posts during the period of data collection, and after a SQL query using the distintic clause, we found that only 72 members performed some type of posting on the supply of services or products. Then in the table client_product were stored information in order to classify what and how often each member posted a particular type of product or service. As shown in Table 2, column actor_id IDs are numbers group members. The columns identified with the initial "p" represent some kind of service or product. Due to the large amount of information from the original table, this table has a partial representation of tabular data. Thus, the 5847576XX member who posted 14 times the product or service offerings and the p2 6295014XX member posted 10 offers the product or service p1, 12 of the product or service p3 and 15 offers the product or service p6.

Given the data collected, together with much information of the posts of digital Tocantins, data from client_product table were exported to a file in csv format, later to be used in this work to present a way of exploring the K-means technique and the measure of centrality.

III. CLUSTERING, DATA PROCESSING AND ANALYSIS

Here are presented the computational results obtained by application of Simple K-means and by betweenness intermediation calculus technique to the extracted from Tocantins Digital Facebook's group and preprocessed as described on section 2.

A. Computational Results Obtained by Using Simple K-means Technique

Simple K-means was considered in this work given that is the most used algorithm for clustering, is of simple implementation and provide good results for the analysed problem.

Simple K-means was the clustering algorithm used to Tocantins Digital facebook's group to find groups of products or services posted by group's members. This way, this technique might be used to vertex positioning that corresponds to product or service names posted in the group

Initially, data from client_product table were exported to a file in csv format to be analyzed in a program written in Java, which used the WEKA libraries, which have the Simple k-means algorithm. The WEKA is a free tool, which has a cluster analysis option and their libraries are available on the developer's site.

After conducting test with different numbers of groups, it was noticed that with grouping of 5 clusters yielded better results due to present the squared error in the amount of 3.74. From the results of squared errors groups with 6, 7, 8, 9 and 10, there is a minimal difference between them. Importantly, clusters with 10 clusters, with the lowest value of squared error, is not ideal, because it increases the possibility of generating multiple clusters with only one member per cluster.

For the composition of the clusters, the algorithm has defined four data sets formed by a standard, which is based on sets that were formed in accordance with range of products/services. To facilitate the viewing of these patterns, as illustrated in Figure 1, were related the following sets:

- The set of P1 to P60, corresponds to computer products, vehicles and telephones/cell phones.
- Set B of P61 to P120, corresponds to appliances and some types of appliance with refrigerator/cooler sub-categories, washing machine and abs.
- Set C, P121 to P180, corresponds to the appliances with the sub-categories air-conditioning/heater and stove; and services as a bricklayer providers, painter, driver, electrician, plumber, cook, housekeeper, day laborer, carpenter, locksmith, husband rentals, realtor, photographer, computer maintenance technician and consultant.
- Set D, P181 to P240, corresponds to the furniture and decorations as bed and mattress, chairs, armchairs and sofas, tables, chests of drawers, shelves and cabinets, and appliances with the cooling sub-categories, dishwasher, oven, clothes dryer / centrifuge and microwave.

From this, the k-means algorithm Simple divided the group of posts in 5 distinct clusters, described as 0 to 4 groups, resulting as follows:

- cluster 0, with 1.5% of the members who have posted products or services that make up the sets A and B.

- cluster 1, with 44% of members who have posted products or services that make up the set A.
- cluster 2, with 36% of the members who have posted products or services that make up the sets B and D.
- Cluster 3, with 17% of the members who have posted products or services that make up the sets A and D.
- cluster 4, with 1.5% of the members who have posted products or services that make up the sets A, B and D.

For this analysis, shown in Figure 1, a trend was detected in the posts of the digital Tocantins members for the products or services of groups 1, 2 and 3. The results of clusters 0 and 4 showed a very low rate of only 1.5 % of members, the group 4 with a member who has published more products or services in clusters a, B and D. Overall C no publications relating to products or services of this subgroup.

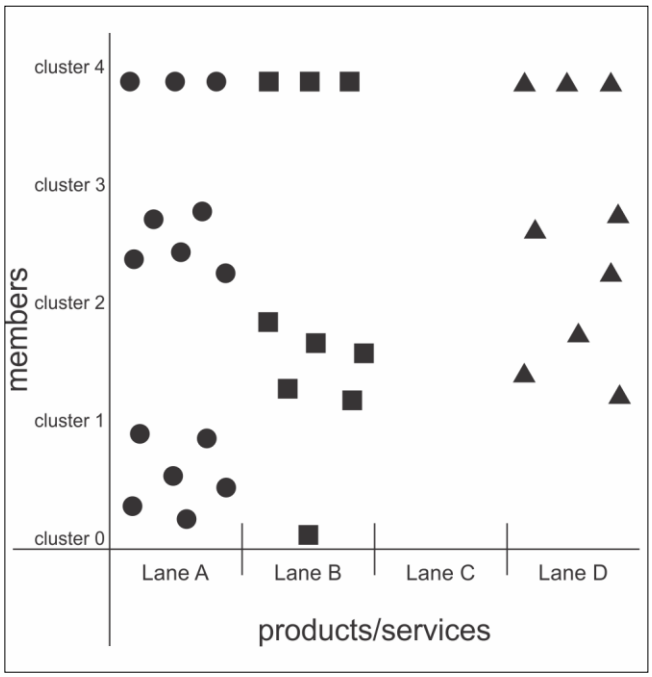


Figure 1. Representation of clusters.

In order to allow a more detailed analysis of the clusters obtained, is outlined in Figure 2, in Venn diagram form, the relationship between the goods / services posted by members of the subset Digital Tocantins. Thus, in Figure 2, one can draw the following analysis. The products / services to set the highest number of publications, with 64% of the total (composite members of clusters 0, 1, 3 and 4). Have Products / D assembly services are in second place with 54.5% of the publications (composed of members of the clusters 2, 3 and 4). The products / services of the set B are third with 39% of publications (composed of members of clusters 0, 2 and 4). Finally, products / C set of services did not have any publication. Thus, it is possible to observe that the posts of computer products, vehicles and / cell phones, which have been put together, have the greatest influence on other products / services from other groups. Therefore, if anyone is interested in any product listed in set A, has a great chance to meet him in Tocantins Digital group.

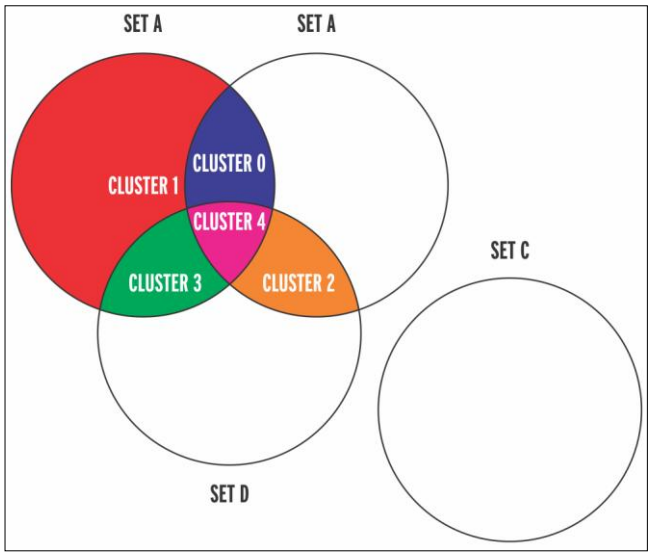


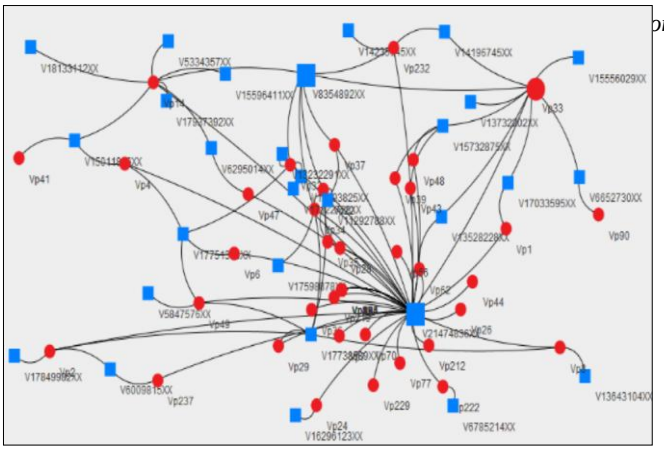
Figure 2. Venn diagram depicting the relationship between Digital Tocantins group members postings divided into clusters by sets of products/services.

B. Computational Results Obtained by Centralization technique application Intermediation

To quantify the graph edges that are considered the most important among members and between members and products, the measure centrality was used. Thus, it is possible to highlight which the vertex is more connected, and this vertex can be a subgroup member that performs more posts, as well as a product or service that are viewed most.

For the beginning of this analysis, data from client_product table were used to identify the vertices and distances (connections) between the vertices. The construction of the graph is done by using the following variables: Distinct_Vertex (indicating all vertices of the graph), Source_Vertex (indicating the origin of a corner connection) Target_Vertex (indicates a connection destination vertex) and Edge_Weight (value edge between two vertices).

The vertices were added by Distinct_Vertex variable to identify 72 vertices corresponding to the actors (members) and 240 corresponding to the vertices posts about products or services. As a result we obtained 110 edges. The zero points of the actors who did not have some kind of post related to products or services have been removed, as well as those products or services that at no time was offered the posts of these members.



After identifying the vertices, the vertices of the links to identify variables were added: The source vertex (Source_Vertex), destination vertex (Target_Vertex) and edge spanning the distance between the vertices (Edge_Weight). These variables were passed as parameters to the BetweenNess_Centrality_Score function, which is responsible for the construction of the type UndirectedSparseGraph graph. Then the calculation is performed by the centrality BetweennessCentrality function, which is part of the package edu.uci.ics.jung.algorithms.scoring JUNG library (Java Universal Network / Graph Framework), available in version 2.0 the manufacturer's website (http://jung.sourceforge.net/). The edu.uci.ics.jung.algorithms.scoring package has mechanisms for the allocation of values, which denote importance, influence, centrality, etc. for elements based on topological properties of the graph [26]. Through VisualizationViewer display, available from JUNG library, along with the Java API JFrame was possible to draw the graph.

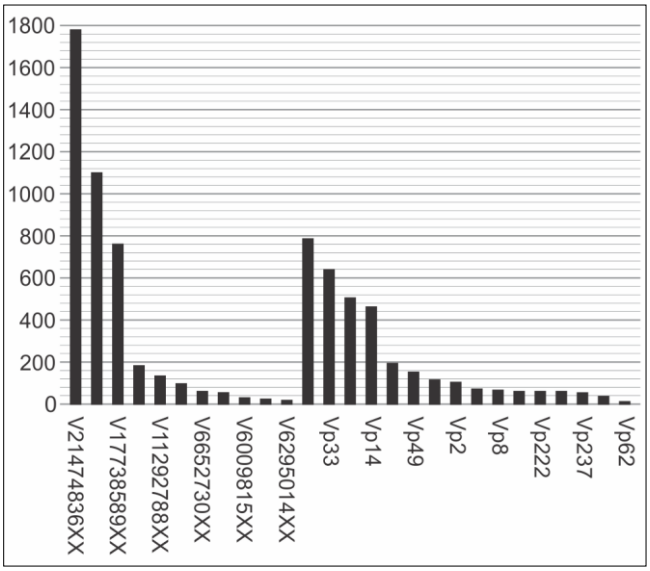


Figure 3. Picture of the most influential vertices of the graph.

In the graph it can be seen in Figure 3, the colors of the vertices of the square shape are represented by the actors (members) and the ellipse format, products or services. The more edges come from a source vertex in square format, more publications are being posted by this actor. The more edges come from a source vertex in the ellipse shape, more often a type of product or service is posted by several members. Figure 3 also illustrates the V21474836XX vertex with higher centrality.

Figure 4. Intermediation centrality measure results.

The result of the algorithm that calculates the mediation centrality measure is shown in Figure 4, with the vertices identified by the initials "Vp" corresponds to any product or service posted in the classified digital Tocantins subgroup. The remaining vertices are the members identified by the numbers of their IDs.

Evaluate the centrality of products and services in the subgroup, offers important information in decision-making for companies in this sector. The vertices represented by users have a greater capacity on their relationships and tend to be better positioned within a group. Stand out of this subgroup analyzed the V21474836XX vertices, V8354892XX and V17738589XX corresponding to subgroup members who have greater centrality. The remaining vertices that represent products and services that have greater centrality are: VP22, Vp33, Vp14 and VP29. The vertices that have not been listed in Figure 4 have mediation centrality values equal to zero.

The information is originated by the posts of the members that form the set of transactions to establish the relationships between the vertices. Thus the intermediation measure on the vertices of the Tocantins Digital subgroup revealed that there are vertices in a better position to transmit and receive information. The V21474836XX vertices, V8354892XX and V17738589XX represent the most influential members of the subgroup, which perform more posts. The VP22 vertices, Vp33, Vp14 VP29 and match the products or services of viewed by members to classifieds listings.

IV. CONCLUSIONS

The works carried out on social networking environment has a large amount of data and thus the use of data mining processes applying graphs, considering that are best represented by vertices corresponding to objects and their edges which form relationships. The dissemination of research work in the area of social networks demonstrate the great value that the information in this environment has. It is also growing forms of data extraction which are provided by the very environment of the social network.

The objectives of this study were achieved with the identification of a number of groups formed by the similarity of products and services. In order to optimize the problem of grouping in search of the best set of center groups with the k-means algorithm, it was realized that the smaller the number of clusters, the greater the similarity between the vertices. From

this, the five groups found it possible to observe the trend of the main products and services offered the posts held by the digital Tocantins members.

Then computed the parameters of centrality, it was possible to build a network topology in order to study the behavior of digital Tocantins subgroup members based on concerns that its members have to view the posts of classified. The view through the graph allowed graphically highlight the behavior of the vertices represented by members of the digital Tocantins subgroup and the vertices represented by most viewed products and services.

With the organization of the data, it was possible to emphasize the importance of research in an online social networking environment, to provide an automated means way in which managers get more information for planning a advertising. Note also that the interactions between members are generated only when a product or service is advertised in the post. By identifying which products and services are most popular and interactions with increased flow between the vertices, you can optimize an advertising system to members of the social network, which are now known through their interests.

Members of a social network are potential future consumers. These consumers tend to search for products and services to meet your specific personal. Therefore, it is interesting to note that currently there is a new setting in the consumption process, mediated by hyper, resulting in consumers who spend also be suppliers of some kind of product or service, that is, can be fostered by individuals or groups with power leadership, linking people around the same interests. The availability and the ease of doing business on social networks show a new e-commerce opportunity due to its large audience among participating members.

Currently there are no more isolated from consumer groups, but a communication network, where people come in contact with each other through social networking sharing opinions through rapid communication. Frequent and repetitive advertising without customer selection criteria will undergo an analysis according to consumer behavior in order to reach only those interested in products or services of interest. Excessive advertising is unnecessary because waste time who discloses randomly and bother consumers who do not wish that offer.

Importantly, the understanding of a network flow behavior constitutes a relevant contribution to managers seeking advertising, ways to reach directly the interests of consumers and provides a way to optimize the profiles analysis of alternatives among people participating closed or open social networking groups.

These results provide managers a way to be updated about the preferences that users of social networks have about the products offered. This shows that managers can not be limited to the company's own website to know the opinion and even to publicize their products and services. Social networks allow a direct opening with consumers because companies can create their own social networking page to be near the customer, enabling data mining achievement in other groups already established.

Social network analysis tools can reveal, through offerings of products and services, both in relation to informal professional, for the companies, the possibility of exchanging experiences between other groups of the same social network searched, as well as a expansion of disclosure.

In social network analysis field, the graph partitioning served as a tool for analysis, modeling, prediction and evolution of these networks, applied to business branches, market analysis, marketing, network infrastructure, relationships, communication between other.

A. Future Works

It is possible to continue in the study communities, applying metaheuristics and making a comparative analysis with the results of k-means for solving problems groupings. Similarly it is also proposed to use the knowledge discovery process (KDD) using the k-means algorithm and the centrality in other social networks to detect their communities.

The techniques discussed in this paper, can serve in research papers seeking preference patterns of products and services in trade and retail sectors but also those simple users who sporadically proffer some kind of product.

REFERENCES

[1] Wasserman, S.; Faust, K. Social network analysis: Method and applications. Cambridge: Cambridge University Press, 1994

[2] Paraiso, G. J. B. O E-Commerce nas rede sociais: estudo sobre os desdobramentos do comércio eletrônico na atualidade. In: Biblioteca on-line de Ciências da Comunicação. 2011. Available in: < http://bocc.unisinos.br/pag/paraiso-gustavo-o-e-comerce-nas-redes-sociais.pdf>. Access in: 21 nov. 2014

[3] Al-Fayoumi, M.; Banerjee Jr, S.; Mahanti, P. K. Analysis of Social Network Using Clever Ant Colony Metaphor. Proceedings of World Academy of Science, Engineering and Technology, v. 41, mai. 2009. Available in <http://mmujallid.kau.edu.sa/Files/611/Researches/54202_24612.pdf> Access in: 27 dez. 2013.

[4] Kirkpatrick, D. Facebook's plan to hook up the world. CNN Money, 29 maio 2007. Available in: < http://money.cnn.com/2007/05/24/technology/facebook.fortune/ >. Access in: 24 dez. 2013.

[5] Patrício, M. R. Tecnologias Web 2.0: Recursos Pedagógicos na Formação Inicial de Professores. Dissertação (Mestrado em Multimédia) Faculdade de Engenharia da Universidade do Porto, Porto, 2009. Available in: < http://hdl.handle.net/10198/1971>. Access in: 29 jan. 2014

[6] Facebook. Developers: The Graph API. Available in: http://www.facebook.com. Access in: 04 mar. 2013.

[7] Lopes, T. J. P. L.; Hiratani, G. K. L. Mineração de opiniões e fatos aplicada à análise de investimentos. São Paulo Senac, 2008.

[8] Tan, Pang-Ning; Steinbach, M.; Kumar, V. Introdução ao Datamining Mineração de Dados. Rio de Janeiro: Editora Ciência Moderna, 2009.

[9] Barabási, A. L. Linked: how everything is connected to everything else and what it means for business, science, and everyday life. Plume, 2003.

[10] Macqueen, J. Some methods for classification and analysis of multivariate observations. 1967. Available in: <www-m9.ma.tum.de/foswiki/pub/WS2010/ CombOptSem/kMeans.pdf>. Access in: 12 dez 2013.

[11] Bishop, C. M. Neural networks for pattern recognition. New York: Oxford University Press, 482p, 1995.

[12] Jain, A. K.; Duin, R. P. W.; Mao, J. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.

[13] Everitt, B. Cluster Analysis. Heinemann Educational Books, London, 1974.

[14] Freeman, L. C. A set of measures of centrality based on betweenness. Sociometry, v. 40, n.1, 1977.

[15] Newman, M. E. J. Fast algorithm for detecting community structure in networks. Physical Review E, 69:026113, 2004.

[16] Mika, P. Social Networks and the Semantic Web. Springer, 2007.

[17] Ferreira, L. N. Técnica de agrupamento de dados baseada em redes complexas para o posicionamento de cluster heads em rede de sensores sem fio.2012. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional). Instituto de Ciências e Matemáticas e de Computação – ICMC-USP. 2012.

[18] Barabási, A. L. et al. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, v. 311, p. 590-614, 2002. Available in: < http://arxiv.org/pdf/condmat/0104162.pdf>. Access in: 23 dez. 2013.

[19] Newman, M. E. J. Networks: an Introduction. Oxford University Press, 2010.

[20] Ströele, V.; Zimbrão, G.; Souza, J. M. Análise de redes sociais científicas: modelagem multi-relacional. In: I Brazilian Workshop Soc Netw Anal Min. 2012.

[21] Ballve, M. Our List of the World's Largest Social Network Shows How Video, Messages, and China are taking over the Social Web. In: Business Insider, 2013. Available in: http://www.businessinsider.com/the-worlds-largest-social-networks-2013-12#ixzz3DVCipYfW. Access in: 23 jan. 2014.

[22] Braga, L. P. V. Introdução à Mineração de Dados. 2. ed. Rio de Janeiro: E-papers, 2005.

[23] Soman, K. P.; Diwakar, Shyam; Ajay, V. Data Mining: Theory and Practice [WITH CD]. PHI Learning Pvt. Ltd., 2006.

[24] Pisani, F.; Piotet D. Como a Web transforma o mundo: a alquimia das multidões. São Paulo: Editora Senac São Paulo, 2010.

[25] Goodrich, M. T.; Tamassia, R. Estruturas de dados e algoritmos em Java. 4 ed. Porto Alegre: Bookman, 2007.

[26] JUNG Java Universal Network/Graph Framework. Libraries. 2009. Available in: <http://jung.sourceforge.net/download.html >. Access in: 04 mar. 2014.