Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

https://doi.org/10.23913/ride.v10i20.603

Artículos Científicos

Análisis de la producción de cuerpos académicos basado en teoría de grafos

Analysis of the Production of Academic Groups Based on Graph Theory

Análise da produção de órgãos acadêmicos com base na teoria dos grafos

Víctor Hugo Menéndez Domínguez
Universidad Autónoma de Yucatán, México
mdoming@correo.uady.mx
https://orcid.org/0000-0003-3587-1263

Jared David Tadeo Guerrero Sosa
Universidad Autónoma de Yucatán, México
jaredgs93@gmail.com
http://orcid.org/0000-0001-7999-9870

María Enriqueta Castellanos Bolaños
Universidad Autónoma de Yucatán, México
enriqueta.c@correo.uady.mx
https://orcid.org/0000-0001-6294-5948

Esmeralda Zurita Gallegos
Instituto Tecnológico Superior de Felipe Carrillo Puerto, México
141k0057@itscarrillopuerto.edu.mx
https://orcid.org/0000-0002-3450-9599

Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Resumen

En muchas universidades mexicanas, grupos de profesores conforman cuerpos académicos con la finalidad de generar nuevos conocimientos con base en la colaboración de sus miembros. Estos cuerpos académicos son clasificados siguiendo criterios establecidos por el Programa para el Desarrollo Profesional Docente (Prodep). Sin embargo, su valoración puede resultar subjetiva por la ausencia de herramientas que faciliten dicho proceso. Este trabajo se enfocó en analizar la producción y la colaboración de tres tipos de cuerpos académicos a partir de la información almacenada en la base de datos Scopus, haciendo uso de la visualización de datos, aplicando la teoría de grafos. La metodología utilizada se basa en la minería de datos educativos. Los resultados permitieron observar una correspondencia entre la estructura del grupo y el cumplimiento de la producción solicitada por el Prodep; los elementos más fuertes y más débiles de cada cuerpo académico fueron localizados con base en su participación y nivel de contribución. El aporte de esta investigación es el análisis de la producción y colaboración científica de un cuerpo académico haciendo uso de la teoría de grafos, lo que permite automatizar el proceso de evaluación y, de esta manera, reducir su interpretación subjetiva.

Palabras clave: colaboración científica, cuerpos académicos, producción científica, teoría de grafos, visualización de datos.

Abstract

In many Mexican universities, groups of professors form academic groups with the purpose of generating new knowledge based on the collaboration of its members. These academic groups are classified according to the criteria established by the Program for Professional Development of Teachers (PRODEP), which is based on the relevance of production and the degree of collaboration; however, such evaluation may be subjective due to the absence of tools that facilitate that process. The present work is focused on analyzing the production and collaboration of three types of academic groups from the information stored in the Scopus database using data visualization, specifically, graph theory. The used methodology is based on educational data mining. The results allowed to observe a correspondence between the structure of the group and the fulfillment of the production requested by the PRODEP, locating the strongest and the weakest elements of each academic group based on their participation and level of contribution. Our research value is the

production and scientific collaboration analysis of an academic group, using a data visualization tool, in this case, graph theory, which allows to automate the evaluation process by reducing its subjective interpretation.

Keywords: scientific collaboration, academic groups, scientific production, graph theory, data visualization.

Resumo

Em muitas universidades mexicanas, grupos de professores formam órgãos acadêmicos para gerar novos conhecimentos com base na colaboração de seus membros. Esses órgãos acadêmicos são classificados de acordo com os critérios estabelecidos pelo Programa de Desenvolvimento Profissional de Professores (Prodep). No entanto, sua avaliação pode ser subjetiva devido à ausência de ferramentas que facilitem esse processo. Este trabalho teve como objetivo analisar a produção e colaboração de três tipos de órgãos acadêmicos, com base nas informações armazenadas no banco de dados Scopus, utilizando visualização de dados, aplicando a teoria dos grafos. A metodologia utilizada é baseada na mineração de dados educacionais. Os resultados permitiram observar uma correspondência entre a estrutura do grupo e o cumprimento da produção solicitada pelo Prodep; Os elementos mais fortes e fracos de cada corpo acadêmico foram localizados com base em sua participação e nível de contribuição. A contribuição desta pesquisa é a análise da produção e colaboração científica de um corpo acadêmico utilizando a teoria dos grafos, que permite automatizar o processo de avaliação e, assim, reduzir sua interpretação subjetiva.

Palavras-chave: colaboração científica, órgãos acadêmicos, produção científica, teoria dos grafos, visualização de dados.

Fecha Recepción: Septiembre 2019 **Fecha Aceptación:** Enero 2020

Introduction

In the scientific community it is a natural activity the collaboration between two or more members (González and Gómez, 2014) that together apply their knowledge to generate new ones, which can be reflected in products of a scientific, educational, technological or inno -vation. Even in Mexico there has been concern about this activity, which has resulted in a policy that stimulates scientific collaboration in public institutions of higher education through the creation of academic bodies (López, 2010). The simplest and most efficient way to know the development and evaluate an academic body is through their publications together (General Directorate of Higher University Education [Dgesu], 2018). There are studies that have focused on representing scientific collaboration through graph theory: from the construction of the graph and its fundamental results (Newman, 2001a) to the practical application of the shortest path theory, networks weights and the centrality of the graph (Newman, 2001b), as well as the use of a system for the visual exploitation of collaborative networks (Giatsidis, Berberich, Thilikos and Vazirgiannis, 2012) and the analysis of scientific collaboration networks considering social factors, their evolution and topical grouping (Staudt, 2011). In the same way, this field has been studied for specific areas, such as biomedical, through the analysis of the characteristics of scientific production in Latin America, where collaboration networks were plotted to obtain the centrality of each network or graph (Huamaní, González, Curioso and Pacheco, 2012).

This article presents an analysis of the level of scientific production of three different groups of teachers through graph theory. These three groups have a different status: an academic body in training (CAEF), a consolidating academic body (CAEC) and a consolidated academic body (CAC), which are affiliated with a Mexican public university. The graphs obtained function as a visualization tool for the evaluation of the productivity and performance of academic bodies. The results obtained could have multiple applications; for example, measuring the relevance in the production of a specific group of teachers, identifying the degree of collaboration that exists between each academic body or helping to improve their performance, among others.

The rest of the article is organized as follows: in the state of the art the concepts necessary to understand the case study are described. Subsequently, the proposed methodology is presented. In the case of study, the concepts of the state of the art are resumed and, together with the methodology, the procedure and the results of the analysis of the scientific production of the academic

bodies in question are presented. Finally, interesting aspects of the results and conclusions are discussed.

State of knowledge

An academic body is not only a fraction of the scientific community, since it has certain specific characteristics. In general, it is a group of professors-researchers who have one or several lines of study in common and aim at both the application and the generation of new knowledge as a result of working together. In Mexico, academic bodies are classified into three groups and each has specific characteristics for state and related universities, institutes and technological universities (Dgesu, 2018). The characteristics presented below correspond to the state and related universities, because the case study is applied with academic bodies of a university belonging to that group.

- Consolidated academic body (CAC). It is the maximum level that an academic body can reach. Its characteristics are the following: 1) the majority of its members have the maximum academic qualification to generate or apply innovative and independent knowledge (doctorate); 2) they have extensive experience in teaching and human resources training; 3) the majority of its members have the desirable profile defined by the Program for Professional Teacher Development, for the Higher Type (Prodep); 4) high commitment to the institution through collaboration and scientific and academic production; 5) demonstrate an intense academic activity manifested in congresses, seminars, tables, workshops, etc., on a regular and frequent basis, and 6) sustain an intense participation in academic exchange networks (Dgesu, 2018).
- Academic body in consolidation (CAEC). It is the intermediate level at which an academic body can be classified. It is characterized by the following aspects: 1) more than half of its members have a doctorate; 2) they have academic products with recognition due to their good quality, derived from consolidated research lines; 3) at least one third of its members have the desirable profile defined by Prodep; 4) participate jointly in lines of research or innovative application of knowledge; 5) they have extensive experience in teaching and human resources training, and 6) collaborate with other academic bodies (Dgesu, 2018).
- Academic Body in Training (CAEF). As the name implies, it refers to academic bodies that are born from one or more lines of research and are at an early stage. Its characteristics are: 1) the members are identified; 2) at least half of its members have the desirable profile defined by

Prodep; 3) they have defined the lines of generation or application of the knowledge they will cultivate, and 4) they have identified the academic bodies related to the one they propose and of high level to establish contact (Dgesu, 2018).

What is sought is that academic bodies have a real impact on society, so it is necessary that they publicize their work in high impact publications. Each academic body has one or more lines of research and the publications they have generated together can be located in repositories that focus on those areas of research. Some of these repositories are the following: dblp for computer science (dblp team [dblp], 2018), arXiv for physics, mathematics, computer science, quantitative finance and statistics (Cornell University, 2018) and Merlot for educational resources Open (California State University System [CSU], 2018), among others. But higher level institutions focus not only on the research areas mentioned and not all repositories have high impact publications. Then it is necessary to consult a solution that meets both characteristics, and Scopus is the right tool. Scopus is the largest database of citations and abstracts of literature reviewed through the peer method. It contains records of scientific journals, books and conference proceedings in a wide variety of fields of science (Elsevier, 2018). Scopus presents intelligent tools to locate, analyze and visualize research, which facilitate the search for authors and additional information such as institutions and publications. It also has its own application programming interface (API), which allows information retrieval through web services (Elsevier, 2017).

On the other hand, a fundamental element in data research is its visualization. Indeed, it is valid to think that it is something simple to do, but the complexity lies in producing good visual representations. In general, data visualization aims to facilitate the exploration and communication of data (Grus, 2015), as well as for its presentation. Although data visualization has been applied mostly in statistics, it is thanks to computing that more benefits are obtained by using tools that allow access to information intuitively based on data (Chen, Härdle and Unwin, 2007) ; In addition, the user, through observation, is able to understand the useful or relevant information of the aspect to be studied, because the interaction with data provides a higher level of understanding (Aparicio and Costa, 2015). A useful technique for visualization and relationship between data is the graph. In computer science, a graph is a mathematical abstraction, represented as G = (V, E), where V is a set of vertices and E is a set of edges. Thus, graphs are useful for modeling relationships between elements and allows the resolution of problems associated with the context in which they are located, and requires a less expensive process than even linear programming. In addition, to

represent a graph there are several options: graphical representation, representation by means of an associated or adjacent matrix, and other representations, such as the graph dictionary (Sallán, Suñé, Fernández and Fonollosa, 2002). As the focus of this article is to represent graphs graphically and analyze their results, it will not deepen the other types of representation. The graphical representation of a graph consists of presenting each vertex as a point or circle, generally. Although it can be represented by other figures, depending on what you want to graph. In the theory of graphs there are interesting concepts that are the object of study. Among them are:

- Degree of a vertex. It is the total number of edges that affect this vertex. It is denoted as g (v) (Álvarez and Parra, 2013).
- Weight. When each edge of a graph G is assigned a real value w (e), it is known as weight (Trudeau, 2017).
- Weighted Graph Graph G along with the weights at their edges (Trudeau, 2017).
- Tagged graph. Graph G has labels either at the vertices, edges or both. When the graph has labels on both, it is said to be a completely labeled graph (Sallán et al., 2002).
- Directed graph. A graph G where it is necessary to represent the origin vertex and the destination vertex, representing the edges by means of arrows (Sallán et al., 2002).
- Non-directed graph. The relationships between the vertices of a graph G are represented by lines (Sallán *et al*., 2002).

The concepts listed are not the only ones; Only those that are useful for the case study of this article are mentioned. There are tools that allow you to create graphs from tables in files with extension .xls and .csv. One of them is NodeXL, which extends to Excel to generate network graphics and graphs (Social Media Research Foundation, 2018). Other alternatives are Gephi (Gephi Consortium, 2018) and Google Fusion Tables (Google, 2018).

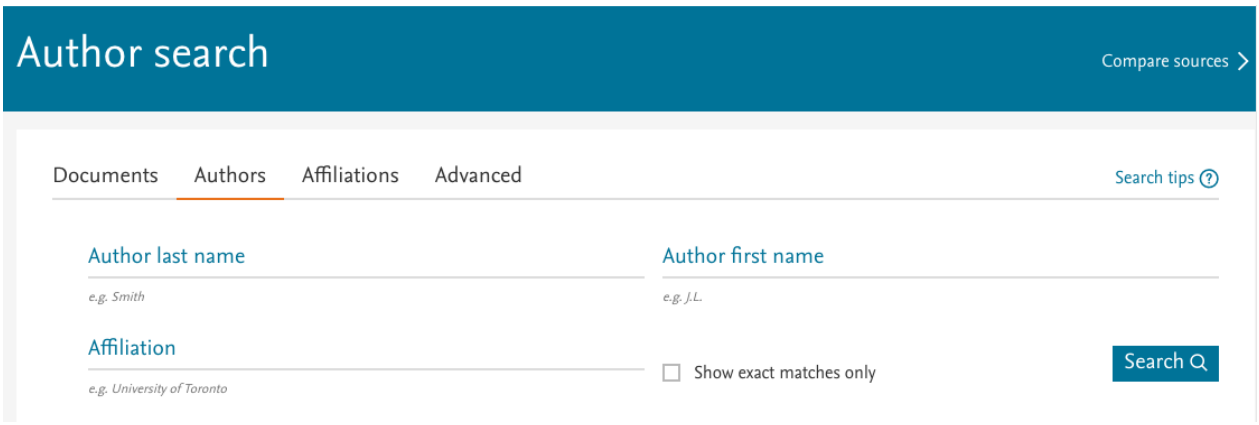
Methodology

Using the methodology of data mining, specifically in e-learning applications (Prieto, Menéndez and Zapata, 2010), a knowledge extraction process is proposed, which consists of the following stages.

Data collection

Once the academic bodies and their members are identified, all the information about their publications is collected using Scopus, since, as mentioned, this tool has data from scientific journals and other high-impact publications, and it is expected that an academic body has them to raise their status, or if consolidated, maintain it. A query is made on the Scopus website for each member of the academic body, entering their name and the institution to which they belong, as seen in the form in Figure 1.

Figura 1. Formulario de búsqueda de perfiles en Scopus



Fuente: Scopus (2019)

Data processing

This stage is divided into four items.

1) Discard duplicate data obtained in the search. When conducting a review in Scopus of each member of the academic body it is logical that, in the case of collaborations, there are publications that belong to two or more members of the academic body.

2) Discard publications that are not related to any other member of the academic body, as well as individual publications.

3) Assign the degree of collaboration of each member of the academic body participating in each publication. For this, the order of the authors in the authorship list is used as a basis, considering that this depends on the degree of contribution of each author in the publication, the first author being the one who contributed the most, and so on. Researchers outside the academic body are not considered. For example, if five authors participate in an article, of

which two belong to the academic body, and the first, third and fourth authors do not, the second author will be considered as the first author and the fifth as the second (for being members of the body academic).

4) Calculate the degree and percentage of contribution per member in the publications.

The list generated in step 3 will be called the Authorship List of the academic body. The degree of collaboration is obtained through the following formula:

$$Grado\ de\ colaboración = 2 * 2^{-p}$$

Where p is the position of the member of the academic body in the list of authorship of the academic body in the publication. What is sought with the degree of collaboration is that its value for the first author of the academic body is 1; of the second author, 0.5; of the third author, 0.25, and so on. In Figure 2 an academic body is presented along with its members as an example, as well as an article called Article 1 and the list of authorship that gives rise to the List of authorship of the academic body, in which the students have been removed authors not belonging to the academic body. The degree of collaboration of the members participating in Article 1 is also presented.

Figura 2. Ejemplo de una publicación perteneciente a un cuerpo académico

Miembros del cuerpo académico
Investigador 1
Investigador 2
Investigador 3

Posición en la lista de autoría en el artículo 1	Autor
1	Externo
2	Investigador 3
3	Externo
4	Externo
5	Investigador 2

Lista de autoría del artículo 1

Posición en la lista de autoría en el artículo 1	Posición en la lista de autoría del cuerpo académico en el artículo 1	Autor	Grado de colaboración
2	1	Investigador 3	1
5	2	Investigador 2	0.5

Lista de autoría del cuerpo académico del artículo 1

Fuente: Elaboración propia

On the other hand, the degree of contribution of each member of the academic body is calculated through the following formula:

$$GCT = \sum_{i=1}^{n} Grado\ de\ colaboración_i$$

As:

- n it is the total of publications where the member in question participates, and
- GCT it is the sum of the weights in the edges that connect the vertex that represents the member of the academic body, and represents the degree of total collaboration of a member to the academic body.

From the value of GTC you can obtain the contribution percentage of an investigator within your academic body following the following formula:

$$PC = \frac{B * GCT}{t}$$

As:

- PC is the percentage of contribution of the member of the academic body (percentage);
- $B = 100$, referring to 100% of the publications of the academic body;
- t is the total number of publications of the academic body, and
- GCT It is the degree of total collaboration of the researcher to the academic body.

Display

With the resulting data, product of the two previous stages, tables are made in NodeXL to generate graphs and carry out the subsequent analysis of the relevant aspects of the academic body studied. Each graph G is represented as:

- Graph $G(V, E)$;
- V is a set of vertices;
- E It is a set of edges;

- VI is a subset of V y VI = { Vertices representing researchers, members of the academic body}, and
- VP It is a subset of V and $VP = \{Vertices\ representing\ academic\ body\ publications\}$.

Likewise, the graph is constructed maintaining the following characteristics:

- The graph is not directed.
- The graph is weighted, and the weights of the edges are the values of the degree of collaboration of each researcher in the academic body involved in each publication.
- The graph is labeled, and the vertices belonging to VI are labeled with the prefix I, then the investigator's identification number is placed. The vertices belonging to VP are labeled with the prefix P and consecutively the identification number of the publication is placed.

Study case

The study that was carried out was the analysis of the production of three academic bodies. Each one belongs to a different group: CAEF, CAEC and CAC. The three academic bodies that were analyzed belong to the Autonomous University of Yucatan (UADY), an important public institution in southeastern Mexico that belongs to the category of state and related universities. Each classified academic body belongs to a different line of research.

Academic body in training

The CAEF analyzed focuses on the diversity and conservation of tropical plant resources, with an impact on the area of natural and exact sciences, specifically the discipline of conservation and management of natural resources. Table 1 shows the degree of the members of the academic body together with their respective production registered in Scopus.

Tabla 1. Datos de los integrantes del CAEF analizado en el caso de estudio

Miembro	Grado	Número de publicaciones en colaboración	Número de publicaciones individuales
Investigador 1	Doctor	16	0
Investigador 2	Doctor	16	0
Investigador 3	Doctor	6	0

Fuente: Elaboración propia

Table 2 shows the publications that relate to CAEF, after debugging external collaborative publications. In this case, there are no individual publications to skip.

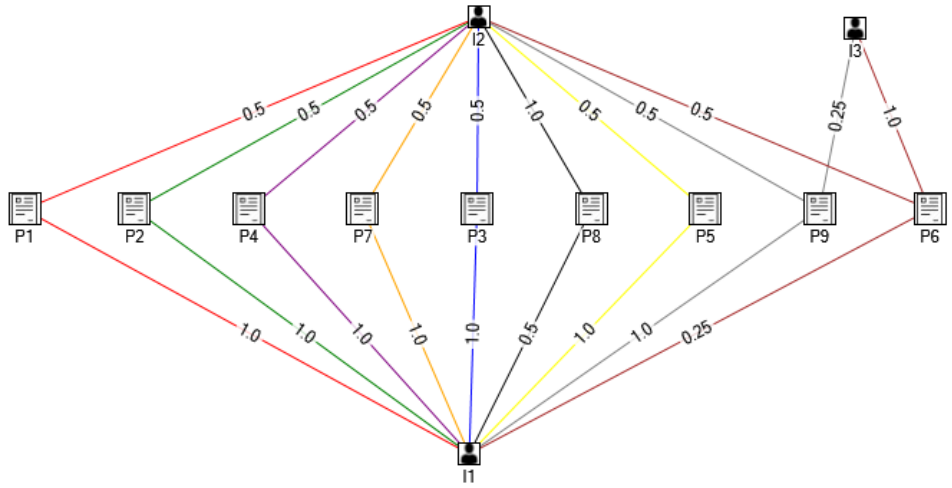
Tabla 2. Publicaciones pertenecientes al CAEF con base en Scopus

Publicaciones	Autores (en orden)
Publicación 1 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 2 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 3 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 4 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 5 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 6 (Artículo)	Investigador 3 (1), Investigador 2 (0.5), Investigador 1 (0.25)
Publicación 7 (Artículo)	Investigador 1 (1), Investigador 2 (0.5)
Publicación 8 (Artículo en prensa)	Investigador 2 (1), Investigador 1 (0.5)
Publicación 9 (Artículo)	Investigador 1 (1), Investigador 2 (0.5), Investigador 3 (0.25)

Fuente: Elaboración propia

Using the results obtained in Table 2 as the basis, a table was created to be processed by NodeXL and build the collaboration graph, which is shown in Figure 3.

Figura 3. Grafo de colaboración entre los miembros del CAEF



Elaboración propia

Results of the CAEF evaluation

Based on the information in table 2 and the graph in figure 2, the analysis of the CAEF's scientific production is carried out according to the type of publication (table 3) and the level of collaboration between members, as well as his contribution in the publications that is reflected in the graph obtained. To carry out the evaluation of the collaboration between CAEF members and their contribution in the publications, the participation of the members was considered, in addition to the degree (ex-putting the total sum of the degree of collaboration in each publication) and the percentage of collaborators. -ration in the publications of the academic body of each of them (table 4).

Tabla 3. Evaluación de la producción científica del CAEF analizado

Tipo de publicación	Total
Artículo en prensa	1
Artículo	8
Producción total	9

Fuente: Elaboración propia

Tabla 4. Evaluación de la colaboración entre los miembros del CAEF y su aporte en las publicaciones con base en el grafo obtenido

Participación de los integrantes (grado de cada vértice) (en orden descendente)	$g(I1)$ = 9 (El Investigador 1 participa en nueve publicaciones) $g(I2)$ = 9 (El Investigador 2 participa en nueve publicaciones) $g(I3)$ = 2 (El Investigador 3 participa en dos publicaciones)
Grado de colaboración total de los investigadores en las publicaciones del cuerpo académico (GCT) (en orden descendente)	Investigador 1: $1 + 1 + 1 + 1 + 1 + 0.25 + 1 + 0.5 + 1 = 7.75$ Investigador 2: $0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 1 + 0.5 = 5$ Investigador 3: $1 + 0.25 = 1.25$
Porcentaje de contribución en las publicaciones del cuerpo académico (PC) (en orden descendente)	Investigador 1: $\frac{100*7.75}{9} = 86.11\ \%$ Investigador 2: $\frac{100*5}{9} = 55.5\ \%$ Investigador 3: $\frac{100*1.25}{9} = 13.88\ \%$

Fuente: Elaboración propia

The most common collaboration that can be inferred from the graph in Figure 2 and the information in Table 4 is between Researcher 1 and Researcher 2 with nine publications in common.

Academic body in consolidation

The CAEC analyzed focuses on research and technological development aimed at generating knowledge for the integration of applications in distributed or parallel computing systems, with an impact in the area of engineering and technology in the discipline of computer science. Table 5 shows the degree of its members with their respective scientific production registered in Scopus.

Tabla 5. Datos de los integrantes del CAEC analizado en el caso de estudio

Miembro	Grado	Número de publicaciones en colaboración	Número de publicaciones individuales
Investigador 4	Doctor	3	0
Investigador 5	Maestro	16	0
Investigador 6	Maestro	10	0
Investigador 7	Doctor	13	0
Investigador 8	Doctor	7	0

Fuente: Elaboración propia

Once the debugging of the publications registered in Scopus that have at least two members of the CAEC, the publications and the corresponding authors are obtained (table 6).

Tabla 6. Publicaciones pertenecientes al CAEC con base en Scopus

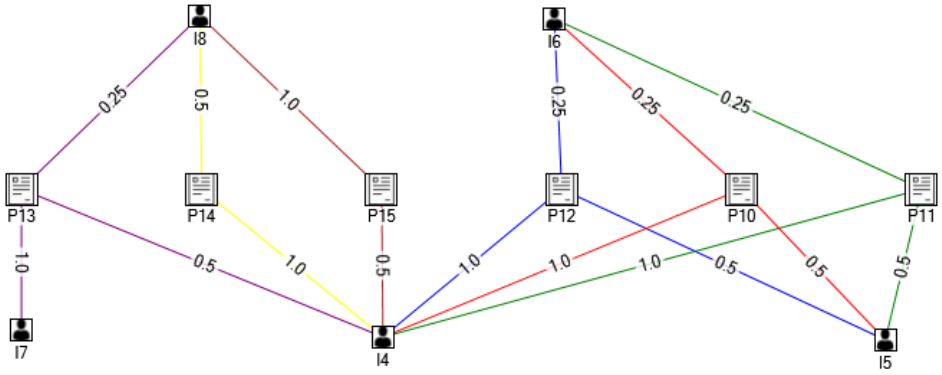
Publicaciones	Autores (en orden)
Publicación 10 (Artículo de conferencia)	Investigador 4 (1), Investigador 5 (0.5), Investigador 6 (0.25) y 2 externos
Publicación 11 (Artículo de conferencia)	Investigador 4 (1), Investigador 5 (0.5), Investigador 6 (0.25) y 2 externos
Publicación 12 (Artículo de conferencia)	Investigador 4 (1), Investigador 5 (0.5), Investigador 6 (0.25) y 2 externos
Publicación 13 (Artículo)	Investigador 7 (1), Investigador 4 (0.5), Investigador 8 (0.25) y 1 externo
Publicación 14 (Artículo)	Investigador 4 (1), Investigador 8 (0.5) y 2 externos
Publicación 15 (Artículo de conferencia)	Investigador 8 (1), Investigador 4 (0.5) y 1 externo

Fuente: Elaboración propia

Using the results obtained in table 6, a table was created to be processed by NodeXL and build the collaboration graph, which is shown in Figure 4.

Revista Iberoamericana para la Investigación y el Desarrollo Educativo
ISSN 2007 - 7467

Figura 4. Grafo de colaboración entre los miembros del CAEC



Fuente: Elaboración propia

Results of the CAEC evaluation

Taking Table 6 and the graph in Figure 4 as a reference, it is evaluated whether CAEC meets the requirements established by Prodep related to scientific production (Table 7). The analysis of the CAEC scientific production analyzed by type of publication (table 8) and the evaluation of the collaboration between the members and their contribution in the publications reflected in the graph obtained are presented in table 9.

Tabla 7. Evaluación de las características generales del CAEC con base en lo establecido por Prodep

Característica	**Cumple**
Cuentan con productos académicos con reconocimiento debido a su buena calidad, derivados de las líneas de investigación consolidadas.	Sí
Participan de forma conjunta en líneas de investigación o aplicación innovadora del conocimiento.	Sí

Fuente: Elaboración propia

Tabla 8. Evaluación de la producción científica del CAEC analizado

Tipo de publicación	Total
Artículo de conferencia	4
Artículo	2
Producción total	6

Fuente: Elaboración propia

Tabla 9. Evaluación de la colaboración entre los miembros del CAEC y su aporte en las publicaciones con base en el grafo obtenido

Participación de los integrantes (grado de cada vértice) (en orden descendente)	$g(I4) = 6$ (El Investigador 4 participa en seis publicaciones) $g(I5) = 3$ (El Investigador 5 participa en tres publicaciones) $g(I6) = 3$ (El Investigador 6 participa en tres publicaciones) $g(I8) = 3$ (El investigador 8 participa en tres publicaciones) $g(I7) = 1$ (El investigador 7 participa en una publicación)
Grado de colaboración total de los investigadores en las publicaciones del cuerpo académico (GCT) (en orden descendente)	Investigador 4: $1 + 1 + 1 + 0.5 + 1 + 0.5 = 5$ Investigador 8: $0.25 + 0.5 + 1 = 1.75$ Investigador 5: $0.5 + 0.5 + 0.5 = 1.5$ Investigador 7: 1 Investigador 6: $0.25 + 0.25 + 0.25 = 0.75$
Porcentaje de contribución en las publicaciones del cuerpo académico (PC) (en orden descendente)	Investigador 4: $\frac{100*5}{6} = 83.3\ \%$ Investigador 8: $\frac{100*1.75}{6} = 29.1\ \%$ Investigador 5: $\frac{100*1.5}{6} = 25\ \%$ Investigador 7: $\frac{100*1}{6} = 16.6\ \%$ Investigador 6: $\frac{100*0.75}{6} = 12.5\ \%$

Fuente: Elaboración propia

The most representative collaborations observed in the graph (figure 4) and in table 9 are:

- Group 1: Researcher 4, Researcher 5, Researcher 6 (three publications).
- Group 2: Researcher 4, Researcher 8 (three publications).

Consolidated academic body

The CAC analyzed focuses on the study of parasites, bacteria and viruses that are the cause of disease in humans, so their work has a greater impact in the area of health sciences, in the discipline of biomedicine . Table 10 shows the degree of its members, as well as their scientific production registered in Scopus.

Tabla 10. Datos de los integrantes del CAC analizado en el caso de estudio

Miembro	Grado	Número de publicaciones en colaboración	Número de publicaciones individuales
Investigador 9	Maestro	37	0
Investigador 10	Doctor	60	0
Investigador 11	Doctor	18	0
Investigador 12	Doctor	9	0
Investigador 13	Doctor	12	0
Investigador 14	Doctor	19	0
Investigador 15	Doctor	7	0
Investigador 16	Doctor	32	0

Fuente: Elaboración propia

Once the debugging of the publications registered in Scopus that have at least two members of the CAC, the publications and the corresponding authors are obtained (table 11).

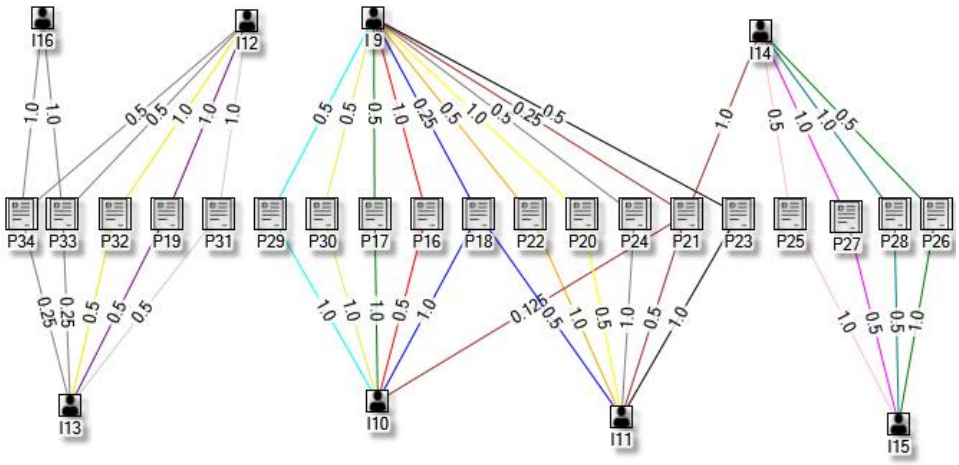
Tabla 11. Publicaciones pertenecientes al CAC con base en Scopus

Publicaciones	Autores (en orden)
Publicación 16 (Artículo)	Investigador 9 (1), Investigador 10 (0.5) y 11 externos
Publicación 17 (Artículo)	Investigador 10 (1), Investigador 9 (0.5) y 5 externos
Publicación 18 (Artículo)	Investigador 10 (1), Investigador 11 (0.5), Investigador 9 (0.25) y 7 externos
Publicación 19 (Nota)	Investigador 12 (1), Investigador 13 (0.5)
Publicación 20 (Artículo)	Investigador 9 (1), Investigador 11 (0.5) y 10 externos
Publicación 21 (Artículo)	Investigador 14 (1), Investigador 11 (0.5), Investigador 9 (0.25), Investigador 10 (0.125) y 9 externos
Publicación 22 (Artículo)	Investigador 11 (1), Investigador 9 (0.5) y 7 externos
Publicación 23 (Artículo)	Investigador 11 (1), Investigador 9 (0.5)
Publicación 24 (Artículo)	Investigador 11 (1), Investigador 9 (0.5)
Publicación 25 (Artículo)	Investigador 15 (1), investigador 14 (0.5)
Publicación 26 (Artículo)	Investigador 15 (1), investigador 14 (0.5)
Publicación 27 (Artículo)	Investigador 14 (1), investigador 15 (0.5)
Publicación 28 (Artículo)	Investigador 14 (1), investigador 15 (0.5)
Publicación 29 (Artículo)	Investigador 10 (1), Investigador 9 (0.5)
Publicación 30 (Artículo)	Investigador 10 (1), Investigador 9 (0.5)
Publicación 31 (Artículo)	Investigador 12 (1), Investigador 13 (0.5)
Publicación 32 (Artículo)	Investigador 12 (1), Investigador 13 (0.5)
Publicación 33 (Artículo)	Investigador 16 (1), Investigador 12 (0.5), Investigador 13 (0.25)
Publicación 34 (Artículo)	Investigador 16 (1), Investigador 12 (0.5), Investigador 13 (0.25)

Fuente: Elaboración propia

Using the results obtained in table 11, a table was created to be processed by NodeXL and build the collaboration graph, which is shown in Figure 5.

Figura 5. Grafo de colaboración entre los miembros del CAC



Fuente: Elaboración propia

Results of the CAC evaluation

Taking Table 11 and the graph in Figure 5 as a reference, and in the same way as in the previous evaluation, it is verified whether the CAC complies with the requirements established by Prodep related to scientific production (Table 12). The analysis of the scientific production of the CAC analyzed by type of publication (table 13) and the evaluation of the collaboration between the members and their contribution in the publications reflected in the graph obtained are presented in table 14.

Tabla 12. Evaluación de las características generales del CAC con base en lo establecido por Prodep

Característica	Cumple
Alto compromiso con la institución por medio de la colaboración y producción científica y académica.	Sí
Demuestran una intensa actividad académica manifiesta en congresos, seminarios, mesas, talleres, etc., de forma regular y frecuente.	Sí
Sostienen una intensa participación en redes de intercambio académico.	Sí

Fuente: Elaboración propia

Tabla 13. Evaluación de la producción científica del CAC analizado

Tipo de publicación	Total
Artículo	18
Nota	1
Producción total	19

Fuente: Elaboración propia

Tabla 14. Evaluación de la colaboración entre los miembros del CAC y su aporte en las publicaciones con base en el grafo obtenido

Participación de los integrantes (grado de cada vértice) (en orden descendente)	$g(I9) = 10$ (El Investigador 9 participa en 10 publicaciones) $g(I10) = 6$ (El Investigador 10 participa en seis publicaciones) $g(I11) = 6$ (El investigador 11 participa en seis publicaciones) $g(I12) = 5$ (El investigador 12 participa en cinco publicaciones) $g(I13) = 5$ (El investigador 13 participa en cinco publicaciones) $g(I14) = 5$ (El investigador 14 participa en cinco publicaciones) $g(I15) = 4$ (El investigador 15 participa en cuatro publicaciones) $g(I16) = 2$ (El investigador 16 participa en dos publicaciones)
Grado de colaboración total de los investigadores en las publicaciones del cuerpo académico (GCT) (en orden descendente)	Investigador 9: $1 + 0.5 + 0.25 + 1 + 0.25 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 5.5$ Investigador 10: $0.5 + 1 + 1 + 0.125 + 1 + 1 = 4.625$ Investigador 11: $0.5 + 0.5 + 0.5 + 1 + 1 + 1 = 4.5$ Investigador 12: $1 + 1 + 1 + 0.5 + 0.5 = 4$ Investigador 14: $1 + 0.5 + 0.5 + 1 + 1 = 4$ Investigador 15: $1 + 1 + 0.5 + 0.5 = 3$ Investigador 13: $0.5 + 0.5 + 0.5 + 0.25 + 0.25 = 2$ Investigador 16: $1 + 1 = 2$
Porcentaje de contribución en las publicaciones del cuerpo académico (PC) (en orden descendente)	Investigador 9: $\frac{100*5.5}{19} = 28.94\ \%$ Investigador 10: $\frac{100*4.625}{19} = 24.34\ \%$ Investigador 11: $\frac{100*4.5}{19} = 23.68\ \%$ Investigador 12: $\frac{100*4}{19} = 21.05\ \%$ Investigador 14: $\frac{100*4}{19} = 21.05\ \%$ Investigador 15: $\frac{100*3}{19} = 15.78\ \%$ Investigador 13: $\frac{100*2}{19} = 10.52\ \%$ Investigador 16: $\frac{100*2}{19} = 10.52\ \%$

Fuente: Elaboración propia

The most representative collaborations observed in the graph (figure 5) and in table 14 are:

- Group 1: Researcher 9, Researcher 10 (six publications).
- Group 2: Researcher 9, Researcher 11 (six publications).
- Group 3: Researcher 12, Researcher 13 (five publications).
- Group 4: Researcher 14, Researcher 15 (four publications).

Discussion

To collect the scientific production it was necessary to perform a search in Scopus with the name and surname of each member. The number of consultations made by each academic body depends on the number of members; Each query was made for each of the Scopus profiles of the members of the academic body in question. However, there are researchers who have more than one profile in Scopus, because they have registered their name in different formats. For example, in some publications there is a registered name and only the father's last name, in others the initial of his name and the two last names, as well as other formats. The greater the number of profiles of the members of the academic body, the greater the number of consultations because, in theory, each profile is a different researcher, which does not reflect reality. The more queries, the longer the data collection time, which could be avoided if each researcher had only one format to write his name and, thus, generated a single Scopus profile. Even a user may not consider the additional profiles, but only the one with the highest number of publications registered, since there may be a belief that it is the most complete, when the reality is that each Scopus profile has different publications.

On the other hand, graph theory, despite being a seemingly simple technique for data visualization, was useful due to the concepts on which it is based, and even made those concepts equivalent with aspects of the type of academic body. In addition to this, at first glance it was observed that CAEF has a greater number of publications than CAEC. But this may be dependent on the year of registration of the academic body and the line of research to which it corresponds. The CAC had the highest number of publications, which allows us to appreciate the constant collaboration and impact of publications in the scientific field. On the other hand, performing the proposed methodology manually is less efficient compared to a technological solution that can take advantage of the Scopus web service (API), and generate a table so that, from it, it is possible to

obtain the graph of automatic way. This table must comply with the format accepted by the tool to be used.

Conclusions

Graph theory can be used to analyze and represent data, as well as for decision-making based on its results, which depends on the institution to which the academic bodies belong or one that evaluates them. In this case, it was useful to know the production of three types of academic bodies following a simple methodology, based on knowledge extraction. It was observed that the three academic bodies comply with what Prodep asks for, in addition to locating the strong or central elements of the academic bodies regarding their participation and level of contribution, the elements that have the least participation and even those that they have a level of contribution considerably far from the highest but they contribute in most publications. As a future work, it is proposed to carry out a technological solution that allows the manual process that was carried out in this case to be carried out automatically, and with that, to do the work more efficiently, in less time and with greater precision, and including So obtain additional and relevant data based on software specialized in graphs such as Gephi.

Acknowledgment

This work has been developed thanks to the support of the National Council of Science and Technology (Conacyt) through the scholarship with number (CVU/Becario): 853088/630948.

References

Álvarez, M. y Parra, J. (2013). *Teoría de grafos*. Bío-Bío, Chile: Universidad del Bío-Bío.

Aparicio, M. and Costa, C. J. (2015). Data visualization. *Communication Design Quarterly Review*, *3*(1), 7–11. Retrieved from https://doi.org/10.1145/2721882.2721883.

California State University System [CSU]. (2018). Merlot. Retrieved from https://www.merlot.org/merlot/index.htm.

Chen, C., Härdle, W.K. and Unwin, A. (eds.) (2007). *Handbook of Data Visualization*. Alemania: Springer Science & Business Media. Retrieved from https://doi.org/10.1007/978-3-540-33037-0.

Cornell University. (2018). arXiv.org e-Print archive. Retrieved from https://arxiv.org/.

dblp team [dblp]. (2018). dblp: computer science bibliography. Retrieved from https://dblp.uni-trier.de/.

Dirección General de Educación Superior Universitaria [Dgesu]. (2018). Dirección General Educación Superior Universitaria | Inicio. México: Dirección General de Educación Superior Universitaria. Recuperado de http://www.dgesu.ses.sep.gob.mx/PRODEP.htm.

Elsevier. (2017). Elsevier Scopus APIs. Retrieved from https://dev.elsevier.com/sc_apis.html.

Elsevier. (2018). Scopus, la mayor base de datos de bibliografía revisada por pares. Recuperado de https://www.elsevier.com/es-mx/solutions/scopus.

Gephi Consortium. (2018). Gephi - The Open Graph Viz Platform. Retrieved from https://gephi.org/.

Giatsidis, C., Berberich, K., Thilikos, D. M. and Vazirgiannis, M. (2012). Visual Exploration of Collaboration Networks based on Graph Degeneracy. Paper presented at the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, August 12-16, 2012.

González, G. y Gómez, J. (2014). La colaboración científica: principales líneas de investigación y retos de futuro. *Revista Española de Documentación Científica*, *37*(4), 1-15.

Google. (2018). About Fusion Tables - Fusion Tables Help. Retrieved from https://support.google.com/fusiontables/answer/2571232?hl=en.

Grus, J. (2015). *Data Science from Scratch* (1st ed.). United States: O'Reilly Media, Inc.

Huamaní, Ch., González, A. G., Curioso, W. H. y Pacheco, J. (2012). Redes de colaboración y producción científica sudamericana en medicina clínica, ISI Current Contents 2000-2009.

Revista médica de Chile, *140*(4), 466-475.

López, S. (2010). Cuerpos Académicos: Factores de Integración y Producción de Conocimiento. *Revista de la Educación Superior*, *39*(155), 7–26.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, *64*(1).

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(1).

Prieto, M., Menéndez, V. y Zapata, A. (2010). Data Mining Learning Objects. En Romero, C., Ventura, S. y Pechenizkiy, M. (coords.), *Handbook of Educational Data Mining* (pp. 315-342). United States: CRC Press.

Sallán, J. M., Suñé, A., Fernández V. y Fonollosa, J. B. (2002). *Métodos cuantitativos en organización industrial I*. España: Edicions UPC.

Scopus. (2019). Search for an author profile. Retrieved from https://www.scopus.com/freelookup/form/author.uri.

Staudt, C. L. (2011). *Analysis of Scientific Collaboration Networks: Social Factors, Evolution, and Topical Clustering*. (diploma thesis). Karlsruhe Institute of Technology, Germany.

Social Media Research Foundation. (2018). NodeXL: Network Overview, Discovery and Exploration for Excel. Retrieved from https://archive.codeplex.com/?p=nodexl.

Trudeau, R. J. (2017). *Introduction to Graph Theory*. Stanford, United States: Stanford University Press.

Rol de Contribución	**Autor (es)**
Conceptualización	Víctor Hugo Menéndez Domínguez
Metodología	María Enriqueta Castellanos Bolaños
Software	Jared David Tadeo Guerrero Sosa
Validación	María Enriqueta Castellanos Bolaños
Análisis Formal	Jared David Tadeo Guerrero Sosa
Investigación	Esmeralda Zurita Gallegos
Recursos	María Enriqueta Castellanos Bolaños
Curación de datos	Esmeralda Zurita Gallegos
Escritura - Preparación del borrador original	Jared David Tadeo Guerrero Sosa
Escritura - Revisión y edición	Víctor Hugo Menéndez Domínguez
Visualización	María Enriqueta Castellanos Bolaños
Supervisión	Víctor Hugo Menéndez Domínguez
Administración de Proyectos	Víctor Hugo Menéndez Domínguez
Adquisición de fondos	Víctor Hugo Menéndez Domínguez