Pesquisa Operacional (2019) 39(3): 405-436
© 2019 Brazilian Operations Research Society
Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
doi: 10.1590/0101-7438.2019.039.03.0405

CENTRALITY METRICS AND GRADUATE PROGRAM EVOLUTION: THE CASE OF PRODUCTION ENGINEERING GRADUATE PROGRAMS IN BRAZIL

Ricardo Lopes de Andrade[1*] and Leandro Chaves Rêgo[2]

Received November 28, 2018 / Accepted September 11, 2019

ABSTRACT. In Brazil, graduate courses are evaluated by the National Graduate Program System and regulated by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), an agency linked to the Ministry of Education (MEC). The intellectual production of the faculty of the programs is the main criteria in determining a program's grade. In this study, we verified whether the grade attributed to the programs is dependent on the co-authorship network of the faculty of the program. Particularly, we analyze whether programs composed mostly by faculty members who cooperate in academic productions and have a more central position in the co-authorship network perform better than those with faculty with fewer collaboration. The paper concludes that there is a relation between the programs' grade and the number of faculty members that collaborates in their intellectual productions. It is also concluded that programs that improved the grade are composed mainly of faculty members with high centrality or have few faculty members with low centrality measures. Moreover, programs that have decreased their grade are formed mainly by faculty members with low centrality measures or with few faculty members with high centrality measures.

Keywords: co-authorship network, graduate programs, evaluation system, social networks.

1 INTRODUCTION

A large part of the scientific publications carried out in Brazil comes from graduate programs. These, moreover, receive resources from governmental funding agencies such as the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), which use, among other tools, the bibliometric data of the researchers to evaluate and distribute their funding to graduate programs

*Corresponding author – https://orcid.org/0000-0001-9639-9797
[1]Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Pernambuco, Avenida da Arquitetura, CEP: 50.740-550, Recife-PE, Brazil
[2]Departamento de Estatística e Matemática Aplicada, Centro de Ciências, Universidade Federal do Ceará, CEP: 60440-554, Fortaleza-CE, Brazil
E-mails: ricardo.lopesa@ufpe.br, leandro@dema.ufc.br

and researchers. The evaluation of graduate programs is an essential activity that aims to ensure and maintain the quality of Masters and PhD courses in the country (CAPES, 2014).

Among the evaluation criteria, Intellectual Production is the most relevant one in defining the programs' classifications. Among the evaluation items of this criterion stands out: "Qualified publications by the faculty of the program" and "Distribution of qualified publications among the faculty members of the program". Journals are classified by CAPES in a ranking, called QUALIS, according to criteria defined by each field of study to evaluate the quality of the scientific production of the programs. In this way, the productivity and quality of the papers published by the faculty of the programs are evaluated.

Studies, such as those by Eaton et al. (1999) and Lee & Bozeman (2005), point to a strong relationship between productivity and collaboration. Hart (2000) research results show that collaboration leads to "improving the quality of publications". Andrade & Rêgo (2015) and Andrade & Rêgo (2017) indirectly also analyzed the effect of collaboration on productivity and quality of publications. They analyzed which Social Network Analysis metrics impact the productivity level of the researchers with CNPq grant in research productivity (PQ) in the area of Industrial Engineering. Cainelli et al. (2015) highlight Specialization, Multi-disciplinarity, Synergy, Opportunity cost of time, Risk diversification, Assigned value of co-authored papers and Social interactions and pressures as determinant factors and facilitators of co-authorship. These results led us to the following research questions: how does collaboration among faculty members impact the Intellectual Production of programs? More broadly, does collaboration impact the programs' classifications?

Scientific research about co-authorship (development of the same paper by two or more authors) has a long history, however, analysis of the co-authorship network to explore scientific collaboration among authors is a relatively new area of research (Uddin et al., 2011). A co-authorship network is a type of social network in which nodes are the researchers and the edges can be formed by the various forms of collaboration existing in the scientific context. The analysis of the patterns, interactions and implications of social networks is facilitated by the use of Graph Theory, through a set of mathematical algorithms (Wasserman et al., 1994). Such an approach is known as Social Network Analysis (SNA). Through this approach, several recent publications have sought to analyze the performance of academic groups based on the structure and dynamics of the relationship among researchers.

The study by de Mello et al. (2010) analyzed the evolution of co-authoring networks among faculty members of Graduate Programs in Management, and the results showed an increase in co-authorship from one triennium to another (2001-2003 and 2004-2006). The results, from Digiampietri et al. (2014) about the Graduate Programs in Computer Science in Brazil and the relationships among them, indicate that programs with higher densities in co-authorship networks also have higher research productivity considering the period from 2004 to 2009. Martins et al. (2014) analyzed the influence of the co-authorship network on the performance of graduate programs in the CAPES evaluation system, concluding that there is a positive relationship among the formation of co-authoring groups and the grades obtained by the programs studied. Andrade

(2016) analyzed the influence of the co-authorship network on the performance of Graduate Programs in Production Engineering in two three-year periods (2007-2009 and 2010-2012) and concluded that the strengthening or development of the Program depends on the co-authorship relations among faculty members of the same program. Amorim (2014) analyzed collaboration networks among doctoral programs in the area of Statistics in Brazil and showed that, simultaneously with the evolution of the networks of the programs, in the period from 2004 to 2012, there was also an evolution in the measures of the networks, as well as the number of collaborations. Abramo et al. (2013) studied the impact of unproductive and top scientists on Italian university performance throught the dataset of all production in 2004-2008 and concluded that top performers impact more than non-productive researchers on university performance.

This paper aims to show how co-authorship networks influenced the development of graduate programs in the area of industrial engineering, called in Brazil as production engineering, in the last three evaluation periods made by CAPES. Specifically, we evaluated the contribution of the major SNA metrics in the evolution of the program classification grades. The remainder of the paper is organized as follows. The next section gives a brief description of the evaluation system of the graduate programs in Brazil. Section 3 recalls the SNA metrics applied in this article. Section 4 describes the methodological procedures to build and analyze the network. Section 5 shows the SNA metrics obtained in the co-authoring networks of the graduate programs and the impact they have on the evolution of their classification grades for three different periods of evaluations. Section 6 sets out the conclusions and closes the article.

2 EVALUATION SYSTEM OF GRADUATE PROGRAMS IN BRAZIL

The evaluation of the National System of Graduate Programs (SNPG) in Brazil, an essential activity to ensure and maintain the quality of Masters and PhD courses in the country, came into force in 1998. This activity is guided by the Evaluation Board/CAPES and carried out with the participation of the academic-scientific community through ad-hoc consultants (CAPES, 2014).

The evaluation system covers two processes (CAPES, 2014):

- Annual Monitoring - carried out in the period between the years of the four-year evaluations. The annual follow-up aims to promote a dialogue between CAPES and the institutions that promote masters and doctoral courses in order to guide the administration of the programs in a form that can raise their quality and to identify possible problems that can be corrected before the subsequent four-year evaluation.

- Four-year Evaluation - is carried out at the end of each quadriennium. The evaluation results of each program are presented in the “Assessment Sheet” defined by the Technical-Scientific Council (CTC). The Assessment Sheet contains the various criteria, items evaluated, the attributes assigned to them and the respective comments and justifications of the organizing committee for giving each attribute. Further, in the end, it is attributed a grade

corresponding to performance of the program in the quadriennium.[1] The grades range from 1 to 7 in increasing order. A course with a grade of less than 3 should be disqualified from the SNPG, while programs that achieve a grade of 3 or greater receive renewal of the "recognition" valid through the subsequent quadrennium. The grades superior to 4 are attributed only to programs that offer doctoral courses, and programs that achieve grades 6 or 7 must represent excellence in the area, recognized as programs of international level.

The criteria evaluated by the SNPG are: Program Proposal (10%); Faculty (20%); Student Body, Thesis and Dissertation (25%); Intellectual Production (35%) and Social Integration (10%). Intellectual Production presents the highest percentage in the programs' grade. It is evaluated in terms of the following attributes: qualified publications by the permanent faculty of the Program (50%); distribution of qualified publications among the permanent faculty members of the Program (30%); technical productions, patents and other production considered relevant (20%). Therefore, it is perceived that scientific production is an important criterion in the evaluation of graduate programs in Brazil.[2]

3 SNA METRICS

Let V be a finite nonempty set and $E \subseteq V \times V$ a set of ordered pairs of elements of V. A graph is a pair $G = (V, E)$, where V is known as the set of nodes of the graph and E is the set of links or edges between the nodes of the graph. A graph can be unweighted if all edges have the same strength or weighted, otherwise. If $n = \#V$, $v_i, v_j \in V$ and the graph is unweighted, the adjacent matrix, A, is an nxn binary matrix whose element, $a(v_i, v_j)$, is defined as follows:

$$a(v_i, v_j) = \begin{cases} 1, & \text{if there is an edge from } v_i \text{ to } v_j, \\ 0, & \text{otherwise.} \end{cases}$$

In the case of a weighted graph, we have that the weighted adjacent matrix of the graph is $W = w(v_i, v_j)_{nxn}$, where $w(v_i, v_j)$ is a non-negative weight of the edge between v_i and v_j. If there is no edge between v_i and v_j, then $w(v_i, v_j) = 0$. Finally, an unweighted graph is called undirected if whenever $a(v_i, v_j) = 1$, then $a(v_j, v_i) = 1$. Otherwise, it is called directed. For a weighted graph to be undirected, it is necessary that for all nodes v_i and v_j, $w(v_i, v_j) = w(v_j, v_i)$. If this is not the case, then it is directed.

A social network can be represented by a graph, where nodes are the actors or individuals in the network and the edges represent some relation between the actors. In particular, a social network of scientific collaboration can be represented by a graph where the nodes are the researchers and

[1] Until 2012, the evaluation occurred in a three-year period. On December 11, 2014, the Higher Council of CAPES decided to change to a four-year evaluation process. The Council also approved a proposal for a mid-term review of a trend analysis of graduate programs.

[2] In December 2018, the Technical-Scientific Council for Higher Education of CAPES, in its 182nd meeting, approved the use of a new Assessment Sheet for the next evaluation, which involves three criteria: Program, Formation and Impact on Society. The weights of each one of the criteria have not been established yet, but scientific production has a significant impact on all criteria.

the edges represent the existence of some joint work between the researchers. The social network of scientific collaboration is unweigthed if the frequency or intensity of collaboration is not taken into consideration and is weighted, otherwise.

In the following subsections, we recall the definitions of some node centrality measures for weighted and unweighted networks. We focus on metrics for undirected networks since scientific collaboration is a symmetric relation.

3.1 Degree Centrality

Proposed by Freeman (1978), the degree centrality (DC) of node v_i, denoted by $C_d(v_i)$, is given by the number of nodes adjacent to vertex v_i. Formally, DC is given by:

$$C_d(v_i) = \sum_{j=1}^{n} a(v_i, v_j). \tag{1}$$

The DC of node v_i, in a weighted network, is given by the sum of all the weights of the edges involving node v_i. For Newman (2004) and Barrat et al. (2014), the weighted DC (WDC) is given by:

$$C_d^w(v_i) = \sum_{j=1}^{n} w(v_i, v_j). \tag{2}$$

The DC is a simple and easy way to measure the local influence of a node (Abbasi et al. (2012); Liu et al. (2005)).

3.2 Closeness Centrality

In order to define closeness centrality, we need to recall what is a path between two nodes v_i and v_j in a network. A path between v_i and v_j is a sequence of nodes $v_0, v_1, \ldots, v_k$, where $v_0 = v_i$, $v_k = v_j$ and $a(v_i, v_{i+1}) = 1$, for $i = 0, 1, \ldots, k-1$. The length of a path is given by the number of nodes in the sequence minus one. In the previous example, the path length is k. The geodesic distance between nodes v_i and v_j, denoted by $d(v_i, v_j)$, is given by the shortest length of all paths from v_i to v_j.

For weighted networks, the weighted length of a path is given by the sum of the reciprocals of the weights of the edges linking consecutive nodes in the path and the weighted geodesic distance, $d^w(v_i, v_j)$, is the minimum weighted length among all paths from v_i to v_j.

The closeness centrality (CC) of node v_i, denoted by $C_c(v_i)$, is given by Freeman (1978):

$$C_c(v_i) = \frac{1}{\sum_j d(v_i, v_j)}. \tag{3}$$

Larger values of $C_c(v_i)$ indicate smaller distances to the other nodes in the network, indicating that the node v_i takes an important position in the network. In weighted networks, the weighted closeness centrality (WCC) is given by:

$$C_c^w(v_i) = \frac{1}{\sum_j d^w(v_i, v_j)}. \tag{4}$$

The closeness centrality of a node measures its independence and efficiency in the communication with other nodes in the network (Freeman, 1978).

3.3 Harmonic Centrality

The harmonic centrality (HC) is a modification of the CC, that is obtained by replacing the average (arithmetic) distance by the harmonic mean of all distances, even for disconnected nodes. The formal definition of the harmonic centrality for unweighted networks, proposed by (Rochat, 2009), is given by:

$$C_H(v_i) = \sum_{i \neq j} \frac{1}{d(v_i, v_j)}. \tag{5}$$

For weighted networks, the definition is given as follows:

$$C_H^w(v_i) = \sum_{i \neq j} \frac{1}{d^w(v_i, v_j)}. \tag{6}$$

3.4 Betweenness Centrality

The betweenness centrality (BC) of vertex v_i, denoted by $C_b(v_i)$, is given by (Freeman (1978); Wasserman et al. (1994)):

$$C_b(v_i) = \sum_{j,k} \frac{g(v_j, v_i, v_k)}{g(v_j, v_k)}. \tag{7}$$

where $g(v_j, v_k)$ is the number of shortest paths between vertices v_j and v_k and $g(v_j, v_i, v_k)$ is the number of shortest paths between vertices v_j and v_k going through vertex v_i.

In a weighted network, the BC is given by:

$$C_b^w(v_i) = \sum_{j,k} \frac{g^w(v_j, v_i, v_k)}{g^w(v_j, v_k)}. \tag{8}$$

where $g^w(v_j, v_k)$ is the number of weighted shortest paths between vertices v_j and v_k and $g^w(v_j, v_i, v_k)$ is the number of weighted shortest paths between vertices v_j and v_k going through vertex v_i, considering the weighted distance, $d^w(v_i, v_j)$.

The greater the BC of a node, the greater the capacity of the node to control the flow of information. According to Freeman (1978); Abbasi et al. (2012), the BC is an indicator of the potential of a node to play the role of "mediator".

3.5 Eigenvector Centrality

A metric of importance of the node in the network based on its connections, the eigenvector centrality is supported on the idea that a particular node will have high centrality if it is connected to vertices with central positions in the network (Bonacich, 1987). In other words, the centrality of the vertex does not depend only on the number of adjacent vertices but also on the centrality of these vertices. Let λ be a constant, then the eigenvector centrality (EC) of node v_i, denoted by $C_e(v_i)$, is given by:

$$C_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} a(v_i, v_j) C_e(v_j). \tag{9}$$

Using the vector notation, let $X = (C_e(v_1), C_e(v_2) \dots C_e(v_n))$ be the vector of eigenvector centralities. We can rewrite Equation 9 as $\lambda X = AX$. By assuming that the eigenvector centrality takes only non-negative values (using the Perron-Frobenius theorem), it can be shown that λ is the largest eigenvalue of the adjacency matrix, where X is the corresponding eigenvector (Jackson, 2008). In the case of weighted networks, the adjacency matrix is replaced by the weighted adjacency (Newman, 2004). And the eigenvector centrality is defined by:

$$C_e^w(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} w(v_i, v_j) C_e^w(v_j). \tag{10}$$

3.6 PageRank

PageRank is one metric used to rank web pages according to the interest and attention devoted to them (Page et al., 1999). PageRank takes into account the number and quality of links to a web page in order to determine how influential it is (Liu et al., 2015). Let T_A be a web page and T_i be one of the web pages that has a link to T_A. Brin & Page (1998) defined PageRank as follows:

$$PR(T_A) = (1-\delta) + \delta \left(\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)} \right), \tag{11}$$

where $PR(*)$ is the PageRank of *, $C(T_i)$ is the number of links leaving page T_i and δ is a damping factor (if a person randomly clicks on pages and eventually stops clicking, then δ is the probability that, at any given moment, the person will continue to click), which can be chosen in the interval $(0,1)$.

3.7 Utility

The study of Rêgo & dos Santos (2019) on co-authorship networks proposed a new SNA metric, specific to this type of network. This metric evaluates the benefit or utility for a given author to be in a certain position in a co-authorship network. The Utility (UT) of node v_i, denoted by $U^w(v_i)$, is given by:

$$U^w(v_i) = \sum_j \left(\frac{w(v_i, v_j)}{C_d^w(v_i)} + \frac{w(v_i, v_j)}{C_d^w(v_j)} + \frac{w(v_i, v_j)^2}{C_d^w(v_i) C_d^w(v_j)} \right), \tag{12}$$

where $w(v_i, v_j)$ is the total number of works involving both authors i and j, $C_d^w(v_i)$ and $C_d^w(v_j)$ are the WDC of these authors, respectively.

The utility described above was based on the original model of the utility of Jackson & Wolinsky (1996). The original metric does not take into account the number of works done between authors. Therefore, the utility of a particular node v_i in an unweighted network is given by:

$$U(v_i) = \sum_j \left(\frac{1}{C_d(v_i)} + \frac{1}{C_d(v_j)} + \frac{1}{C_d(v_i)C_d(v_j)} \right), \tag{13}$$

where $C_d(v_i)$ and $C_d(v_j)$ are the DC of vertices v_i and v_j, respectively.

3.8 Local Clustering Coefficient

The local clustering coefficient (CL) measures the degree of the density of the neighborhood of connections of a given node, i.e., it corresponds to the degree to which the neighbors of a node interconnect (Barabási, 2015). The CL, in a co-authoring network, of a given node indicates how his collaborators are working together (Onel et al., 2011).

The clustering coefficient of node v_i is the ratio between the number of triangles that contains node v_i and the number of possible edges between the neighboring nodes. Let $NT(v_i)$ be the number of triangles (consists of three nodes connected by three links) containing node v_i. For Onnela et al. (2005), the local cluster coefficient is defined as:

$$CL(v_i) = \frac{2NT(v_i)}{C_d(v_i)(C_d(v_i) - 1)}. \tag{14}$$

The weighted local clustering coefficient was proposed by Onnela et al. (2005) and is given by:

$$CL^w(v_i) = \frac{2}{C_d(v_i)(C_d(v_i) - 1)} \sum_{j,k} \left((\hat{w}(v_i, v_j)\hat{w}(v_i, v_k)\hat{w}(v_j, v_k) \right)^{1/3}, \tag{15}$$

where the weights of the edges are normalized by the maximum weight of the network, $\hat{w}^(v_i, v_j) = w(v_i, v_j)/\max_{v_i, v_j \in V}(w(v_i, v_j))$ and the contribution of each triangle depends on all the weights of the edges.

3.9 Summary of measures

Table 1 summarizes the social network measures described in this section.

Table 1 – Measures applied in a co-authorship network analysis

Measures	Unweighted	Weighted	Practical meaning
Degree centrality	DC	WDC	In a co-authorship network, this metric identifies the authors with highest number of collaborators (unweighted) and with highest number of papers in co-authorship (weighted) (Abbasi et al. (2011); Anastasios et al. (2012); Freeman (1978)).
Closeness Centrality	CC	WCC	The closeness centrality of a node measures its independence and efficiency in the communication with other nodes in the network (Freeman, 1978).
Harmonic centrality	HC	WHC	Such as the closeness centrality, the harmonic centrality of a node measures its independence and efficiency in communicating with other nodes in the network.
Betweenness Centrality	BC	WBC	The greater the BC of a node, the greater the capacity of the node to control the flow of information. According to Freeman (1978); Abbasi et al. (2012), the BC is an indicator of the potential of a node to play the role of "mediator".
Eigenvector Centrality	EC	WEC	By eigenvector centrality, researchers who are connected to more central researchers, according to eigenvector centrality, have greater centralities.
PageRank	PR	WPR	The importance or prestige of nodes defined by PageRank, unlike measures of degree, closeness, and betweenness centrality, is modeled or transferred from the importance of adjacent nodes (Liu et al. (2005)). Thus, the prestige of the faculty depends on the prestige of its co-authors.
Utility	UT	WUT	Evaluates the benefit or utility for a particular author of belonging to a certain network structure.
Local Clustering Coefficient	CL	WCL	The clustering coefficient of a particular researcher indicates how much their collaborators are collaborating with each other.

4 METHODOLOGY

The study, characterized as a descriptive research, associates the evolution or development of the performance of the Graduate Programs in Production Engineering (PPGEP) with regard to the grade attributed by CAPES to the principal centrality metrics of the social co-authorship network, where nodes represent all permanent faculty of all the PPGEP and their co-authors. A single network was built for each period of evaluation.

For the construction of the co-authorship network among PPGEP faculty members and their co-authors only journal publications by permanent faculty of some PPGEP were considered. Moreover, only publications in three periods of CAPES evaluation were considered: 2007-2009, 2010-2012 and 2013-2016. The following steps were taken:

- Step 1 - Identification of PPGEP in the last three evaluations by CAPES. The PPGEP of the last three evaluations 2007-2009, 2010-2012 and 2013-2016 were extracted from the List of Recommended and Recognized Courses by CAPES corresponding to the Production Engineering area and that has the nomenclature of Production Engineering. Programs that although related to production engineering have other nomenclatures and extensions, such as Production Engineering and Systems or Industrial Engineering, were excluded from the analysis. In total, 20 programs were selected for the study;

- Step 2 - Identification of permanent faculty, associated to each PPGEP. For each period of evaluation and for each Program, the list of permanent faculty was identified. Altogether, 336 faculty belonging to at least one PPGEP evaluated in the triennium 2007-2009 were identified, in the triennium 2010-2012, this number was 334 and finally in the last evaluation period 2013-2016 the number of permanent faculty raised to 352;

- Step 3 - Identification of Lattes Curricula[3] of each faculty identified in the previous step. Lattes IDs were obtained directly from faculty's Lattes curriculum.

- Step 4 - Extraction of the academic productions of each faculty. The scriptLattes, a free code tool designed for extracting and automatically compiling information contained in the Lattes Curricula (Mena-Chalco & Junior, 2009) was used to extract academic productions. In the text file, input source for the tool, included in each line, in this order and comma-separated: the Lattes ID; the name of the faculty; the period in which the faculty belonged to the Program, in each evaluation period; and the Program that it belongs to as a label. Among several files, the scriptLattes generates a list of publications made in co-authorship, a list of co-authors and co-authorship relationships.

- Step 5 - Calculation of SNA metrics. With the co-authoring file, generated by scriptLattes, weighted and non-weighted metrics were obtained using NetworkX (Hagberg et al., 2008), which consists of a set of open codes developed in Python language (Team, 2015).

[3]Lattes is a platform (lattes.cnpq.br), developed by the CNPq in Brazil, for the database integration of curricula of people working in academic or research institutions.

- Step 6 - Partition of the PPGEP according to whether their grade had increased, decreased or remained constant.

The Chi-square test of association indicates whether two categorical variables are associated or not, i.e., it determines whether the distribution of observations of one variable varies depending on the category of the second variable.

Our first statistical analysis was to verify, by means of a Chi-square test, whether the number of faculty who carried out journal publications with or without co-authorship was associated with the evolution of the PPGEP grade in each period. For that, for each faculty in each period of evaluation, it was verified what was the evolution of the programs to which he belongs and if the faculty had a journal publication in co-authorship or not.[4]

The analysis was divided in two cases: one which only considered a co-authorship when at least two faculty members of the same program were co-authors in the same paper (endogenous co-authorship) and another which only considered a co-authorship when a faculty member of some program published a joint paper with a co-author who was not a faculty member of the same program in the period (exogenous co-authorship).

The following hypotheses were formulated:

- H0: the evolution of the program's grade does not depend on its faculty co-authorship.
- H1: the evolution of the program's grade depends on its faculty co-authorship.

Next we evaluated whether the existed an association between (non-weighted and weighted) SNA metrics and the evolution of the programs' grades.

The weights of the edges in the co-authorship networks were given by the number of joint works between the pair of authors.

In order to use the Chi-square test, we transformed the SNA metrics into categorical variables. Three classes were established: low centrality (equal to or below the 33.333 percentile); high centrality (equal to or higher than the 66.666 percentile) and medium centrality (when not included in the two previous classes).

To perform the Chi-square test, for each faculty in each period of evaluation, it was verified what was the evolution of the programs to which he belongs and in which category (low, medium or high) its centrality measure was. The following hypotheses were formulated:

- H0: the evolution of the program's grade does not depend on the centrality of its permanent faculty.
- H1: the evolution of the program's grade depends on the centrality of its permanent faculty.

[4]In the very few cases, where a single faculty belonged in the same period to more than one PPGEP, he entered into all the analysis as a separate observation for each program that he belonged to.

The chi-square test is based on a hypothesis test that is able to identify the general association between the variables. However, the recognition of particular situations, intersections between categories of variables, is only possible through the residue assessment. Residues are indicators that represent the difference between observed and expected values. When expressed in units of standard deviation they are identified as standardized residues and when they are weighted by the variance they are called adjusted residues.

The adjusted standardized residuals analysis, therefore, is used to complement the chi-square test, and hence indicates whether a particular joint category occurred more or less than what is expected if the variables were independent. Therefore, these values should be compared to the critical points of a standardized curve. When using a 95% confidence level, the existence of positive values greater than 1.96 indicates more occurrences observed than expected, and negative values lower than -1.96 signify the opposite. Hence, the larger the absolute value of the adjusted residue, the greater the association between the categories (Field (2013)).

5 DATA ANALYSIS

We start this section by giving a summary about the programs' grades in the three evaluation periods. Table 2 shows the distribution of program grades in all three periods. One can see that the median grade remained constant and equal to 4 in all three periods, but the range of grades increased. In Table 3, the distribution of program grade evolution is displayed for all three periods. In all three periods, programs that maintained their grade was the majority, and the number of programs which had a decrease in their grade was substantially higher in the last two periods in relation to the first one.

Table 2 – Distribution of program grades in the three evaluation periods.

Grade	2007-2009	2010-2012	2013-2016
2	0	0	2
3	6	5	1
4	8	7	9
5	5	6	6
6	1	2	0
7	0	0	2

Table 3 – Distribution of the evolution of the program grades in the three evaluation periods.

Grade evolution	2007-2009	2010-2012	2013-2016
Grade fell	1	4	4
No change	12	8	10
Grade rose	7	8	6

The following subsections present the results of the chi-square test and residue analysis for each one the analysis. The first analysis is about the association of programs' grades and co-authorship

relations of its permanent faculty. Then each subsection addresses the association of an SNA metric, with or without edge weight, with program's performance. The networks formed by the co-authorship relationship considered the publications in periodicals of the permanent faculty of the PPGEP, through the file obtained by scriptLattes. In the first evaluation period (2007-2009), the network consisted of 1317 nodes, where 336 of these represented permanent faculty in PPGEP. The co-authorship of these faculty was represented by 2187 edges, resulting in a density of 0.0025 and in a clustering coefficient of 0.5489. In the second period (2010-2012), the network consisted of 2036 nodes, where 334 of these represented permanent faculty in PPGEP. The co-authorship of these faculty was represented by 4388 edges, resulting in a density of 0.0022 and in a clustering coefficient of 0.6082. Finally, in the third period (2013-2016), the network consisted of 3078 nodes, where 352 of these represented permanent faculty in PPGEP. The co-authorship of these faculty was represented by 8901 edges, resulting in a density of 0.0019 and in a clustering coefficient of 0.6057.

5.1 Co-authorship Relations

The result of the Pearson chi-square test in Table 4 shows that the program grade performance depends on the publications carried out by the faculty, with or without co-authorship. In other words, we reject H0, at the significance level of 5%. Of the last three evaluation periods, the grade evolution of the triennium 2010-2012 was the one that presented the highest association with co-authorship.

Table 4 – Performance vs. Endogenous Co-authorship - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	6.029	30.638	12.249
df	2	2	2
Asymp. Sig. (2-sided)	0.049	0.000	0.002

Table 5 – Performance vs. Endogenous Co-authorship - Adjusted residual analysis

	2007-2009		2010-2012		2013-2016	
	No	Yes	No	Yes	No	Yes
Grade fell	1.89	-1.89	**3.72**	**-3.72**	**2.74**	**-2.74**
No change	1.07	-1.07	1.94	-1.94	**-3.23**	**3.23**
Grade rose	-1.79	1.79	**-5.34**	**5.34**	1.24	-1.24

In Table 5, the triennium 2007-2009 there was no significant association among the classes of the variables. Hence, although the chi-square test identified a relationship of dependence with marginal significance between the variables, the analysis of the residuals did not identify any individual discrepancy between the categories of the variables in relation to the expected values.

In the following period there is a strong association between cooperating and the grade rise (5.34 of adjusted residue), as well as a strong association between the grade fell and the non-cooperation of the teaching staff (3.72 adjusted residue). The positive influence of the cooperation between the permanent teachers in the grade performance was also found in the last evaluation period. The residual adjustment of 2.74 indicates a strong association between the grade fell and non-cooperation. Also during this period, programs that did not change the grade were positively influenced by the great participation of collaboration among teachers.

Table 6 – Performance vs. Exogenous Co-authorship Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	6.036	8.800	2.370
df	2	2	2
Asymp. Sig. (2-sided)	0.049	0.012	0.306

Table 7 – Performance vs. Exogenous Co-authorship - Adjusted residual analysis

	2007-2009		2010-2012		2013-2016	
	No	Yes	No	Yes	No	Yes
Grade fell	1.71	-1.71	1.22	-1.22	1.31	-1.31
No change	1.30	-1.30	1.84	-1.84	0.08	-0.08
Grade rose	-1.95	1.95	**-2.96**	**2.96**	-1.17	1.17

In Table 6, the result of the Pearson chi-square test shows that the grade program performance was influenced by the collaboration between faculty and external collaborators, except in the last evaluated period. Based on the analysis of residue, Table 7 only in the 2010-2012 period, we observed an excess of teachers who established external co-authoring relationships in programs whose grade rose. Therefore, one way to improve or at least maintain a program's grade is through encouraging academic collaboration.

In the following subsections, we investigate the dependence of the evolution of the grade of a program in the last three evaluation periods and some SNA metrics. Dependence is investigated through the use of Pearson chi-square tests and adjusted residual analysis.

5.2 Degree Centrality

Table 8 shows that there was a dependence between the evolution of a program's grade and the DC category of its permanent faculty in the triennia 2007-2009 and 2010-2012, but not in the quadrennium 2013-2016. To infer how the DC category influenced the programs evolution, we used standardized adjusted residuals. These results are displayed in Table 9.

For the triennium 2007-2009, it can be seen that the presence of permanent faculty with high DC is positively associated with an increase in the program's grade and negatively associated with

a no change in the grade. On the other hand, the presence of permanent faculty with low DC is not associated with a decrease in the grade, but it is positively associated with no change in the grade and negatively associated with an increase in the grade. For the triennium 2010-2012, it can be seen that the presence of permanent faculty with high DC is also positively associated with an increase in the program's grade, but now it is negatively associated with a decrease in the grade. On the other hand, the presence of permanent faculty with low DC is now associated with a decrease in the grade and continuous to be negatively associated with an increase in the grade.

Therefore, in these two periods, programs with more cooperative faculty had a positive inclination in performance. In the final period of evaluation, no dependence was found between the evolution of the program's grade and DC category of its faculty.

Table 8 – Performance vs. DC Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	15.612	20.103	6.180
df	4	4	4
Asymp. Sig. (2-sided)	0.004	0.000	0.186

Table 9 – Performance vs. DC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.93	1.60	-0.72	**-2.55**	**2.66**	-0.20	-1.91	1.55	0.38
No change	**-3.22**	**2.02**	1.20	-1.35	1.56	-0.27	1.03	- 1.89	0.86
Grade rose	**3.63**	**-2.65**	-0.96	**3.68**	**-4.00**	0.45	0.47	0.77	-1.26

Tables 10 and 11 show the results for similar analysis now considering a weighted network. The chi-square test also shows that only in the first two period of evaluations there exists dependency between evolution of the program's grade and the WDC category of its faculty.

Positive association was observed between high WDC and increase in grade in the first two periods of evaluations and between low WDC and decrease in grade in the last two periods. Moreover, negative association was found between low WDC and increase in grade in the first two periods, between high WDC and no change in grade in period 2007-2009, and between high WDC and decrease in grade in period 2010-2012. It can also be observed that the degree of association increased in the last two periods of evaluation when weights are taken into consideration. The opposite was found in the first period of evaluation. Thus, not only the number of collaborators, but also the frequency of the collaboration turns out to be important in the evolution of the program's evaluation.

Table 10 – Performance vs. WDC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	13.452	32.717	9.082
df	4	4	4
Asymp. Sig. (2-sided)	0.009	0.000	0.059

Table 11 – Performance vs. WDC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-1.07	1.51	-0.49	**-3.34**	**3.72**	-0.42	-2.35	**2.82**	-0.55
No change	**-2.98**	1.63	1.45	-1.83	1.14	0.69	1.16	-1.02	-0.12
Grade rose	**3.43**	**-2.22**	-1.30	**4.88**	**-4.54**	-0.31	0.69	-1.23	0.59

These results indicate that either increasing the number of collaborators or the number of papers with existing collaborators tend to improve a program's grade.

5.3 Closeness Centrality

Tables 12 and 13 show that, in all periods of evaluations and considering or not the edges' weights, there exists a dependency between the programs evolution and the (weighted) closeness centrality (CC) category of its permanent faculty. This association was more intense when considering the edges' weights. The closeness centrality metric is used to measure the independence of faculty and those with high CC have higher chance of establishing new publications partnerships.

Table 12 – Performance vs. CC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	13.220	24.016	34.597
df	4	4	4
Asymp. Sig. (2-sided)	0.010	0.000	0.000

Table 13 – Performance vs. WCC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	17.728	30.894	37.536
df	4	4	4
Asymp. Sig. (2-sided)	0.001	0.000	0.000

In the analysis of residues shown in Tables 14 and 15, we can notice that programs that had a decrease in the grade show a higher number of faculty with low CC and WCC (2010-2012 and 2013-2016). The programs that maintained the grade presented, in the first period, a high number of faculty with low CC and WCC, but in the last period, faculty with high CC and WCC were predominant in these programs. Regarding the programs that had an increase in the grade, it is observed that in the first two periods are constituted mainly by faculty with high CC and WCC, while in the last period we can conclude that these programs do not present a significant number of faculty with low CC and WCC.

Table 14 – Performance vs. CC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.56	-0.70	**-3.21**	**2.59**	0.62	**-5.27**	**4.81**	0.47
No change	**-2.59**	**2.42**	0.12	-0.22	1.94	-1.72	**2.41**	-1.66	-0.75
Grade rose	**2.97**	**-3.04**	0.14	**3.15**	**-4.31**	1.17	1.75	**-2.18**	0.43

Table 15 – Performance vs. WCC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.82	-0.91	**-3.48**	**3.92**	-0.44	**-5.27**	**5.10**	0.17
No change	**-3.31**	**2.26**	1.05	-0.70	1.22	-0.52	**2.86**	-1.66	-1.20
Grade rose	**3.72**	**-2.97**	-0.74	**3.87**	**-4.79**	0.92	1.26	**-2.42**	1.17

These results suggest that programs should encourage collaboration of its faculty members, specially establishing new connections with other researchers which are further away in the network to reduce the average distance of the program to all other faculty in the network.

5.4 Harmonic Centrality

As in the case of the closeness centrality, Tables 16 and 17 show that, in all periods of evaluation and considering or not the edges' weights, there exists a dependency between the evolution of the program's grade and the category of the (weighted) harmonic centrality (HC) of its permanent faculty. It was also observed that the association was more intense when considering the edges' weights. Like the closeness centrality, the HC is another metric used to measure the independence of faculty and those with high HC have higher chance of establishing new publications partnerships.

Table 16 – Performance vs. HC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	12.128	29.038	31.761
df	4	4	4
Asymp. Sig. (2-sided)	0.016	0.000	0.000

Table 17 – Performance vs. WHC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	14.348	37.792	35.311
df	4	4	4
Asymp. Sig. (2-sided)	0.006	0.000	0.000

In the analysis of residues shown in Tables 18 and 19, we can notice that programs that had a decrease in the grade show a higher number of faculty with low HC and WHC (2010-2012 and 2013-2016). The programs that maintained the grade presented, in the first period, a high number of faculty with low HC and WHC, but in the last period, faculty with high HC and WHC were predominant. Regarding the programs that had an increase in the grade, it can be observed that in the first two periods, they were constituted mainly by faculty with high HC and WHC, while in the last period, we can conclude that these programs do not present a significant number of faculty with low HC and WHC.

Table 18 – Performance vs. HC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.70	-0.82	**-3.74**	**2.59**	1.15	**-5.27**	**4.21**	1.06
No change	**-2.34**	**2.18**	0.14	0.02	**2.18**	**-2.21**	**2.63**	-1.43	-1.20
Grade rose	**2.73**	**-2.85**	0.15	**3.39**	**-4.55**	1.17	1.50	-1.93	0.43

Table 19 – Performance vs. WHC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.82	-0.91	**-4.01**	**3.39**	0.62	**-4.97**	**5.10**	-0.13
No change	**-2.83**	**2.02**	0.81	-1.42	1.94	-0.52	**2.63**	-1.66	-0.98
Grade rose	**3.22**	**-2.73**	-0.50	**5.08**	**-5.03**	-0.04	1.26	**-2.42**	1.17

As in the case of closeness centrality, the results of harmonic centrality analysis suggest that programs should encourage collaboration of its faculty members, specially establishing new connec-

tions with other researchers which are further away in the network to reduce the average distance of the program to all other faculty in the network.

5.5 Betweenness Centrality

Tables 20 and 21 show an association between the (weighted) betweenness centrality (BC) with the evolution of the performance of the programs in the analyzed periods, except in the quadrennium 2013-2016, when the metric is unweighted. Moreover, in all periods of evaluation, when weights are considered the intensity of the association increases. The BC measures the importance of faculty as mediators, who control the flow of information in the network. Therefore, a faculty with a high level of intermediation is an indicative that most of the relationships established in the network are directly or indirectly linked to his publications.

Table 20 – Performance vs. BC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	16.003	13.707	8.464
df	4	4	4
Asymp. Sig. (2-sided)	0.003	0.008	0.076

Table 21 – Performance vs. WBC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	16.049	20.425	12.077
df	4	4	4
Asymp. Sig. (2-sided)	0.003	0.000	0.017

By the analysis of residues shown in Tables 22 and 23, it can be noted that programs with a high number of faculty with high BC and WBC are positively associated with an increase in the programs' grade in the first two periods of evaluation, is negatively associated with a no change in the grade in the first period of evaluation and is negatively associated with a decrease in programs' evaluation in the period of 2010-2012. Finally, in the last period, a high number of faculty with high WBC is also negatively associated with a decrease in the programs' grade.

On the other hand, in the first period of evaluation, programs with a high number of faculty with low BC and WBC are positively associated with a no change in the programs' grade and negatively associated with an increase in the programs' grade. In the second period of evaluation, a high number of faculty with low BC and WBC is negatively associated with an increase in the programs' grade while only a high number of faculty with high WBC is positively associated with a no change in the programs' grade. Finally, in the last period of evaluation, a high number

of faculty with a low BC and WBC is positively associated with a decrease in the programs' grade.

Table 22 – Performance vs. BC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	0.77	0.12	**-2.16**	1.80	0.36	-1.71	**2.73**	-1.02
No change	**-3.31**	**2.89**	0.35	-0.70	1.70	-1.00	0.60	-0.53	-0.07
Grade rose	**3.72**	**-3.23**	-0.40	**2.67**	**-3.35**	0.68	0.77	-1.69	0.92

Table 23 – Performance vs. WBC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-1.59	1.34	0.21	**-2.95**	1.54	1.42	**-2.01**	**3.32**	-1.32
No change	**-2.83**	**2.74**	-0.01	-0.94	**2.18**	-1.24	0.15	-1.21	1.07
Grade rose	**3.47**	**-3.28**	-0.07	**3.63**	**-3.59**	-0.04	1.50	-1.44	-0.06

To increase the betweeness centrality of its faculty members and, consequently, increasing a program's grade, it is necessary to increase the number of short paths that pass through its faculty members. Again, this can be achieved by encouraging collaboration with the faculty which are further away in the network.

5.6 Eigenvector Centrality

In Tables 24 and 25, the chi-square test indicated an association of (weighted) eigenvalue centrality (EC) with the evolution of the programs' grade, except that the association in the 2007-2009 triennium for the unweighted EC was not significant at the significance level of 0.05. A high value of EC indicates that the faculty has direct collaboration with other influential faculty in the network.

Table 24 – Performance vs. EC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	9.233	34.530	25.623
df	4	4	4
Asymp. Sig. (2-sided)	0.056	0.000	0.000

Table 25 – Performance vs. WEC - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	12.587	35.014	35.755
df	4	4	4
Asymp. Sig. (2-sided)	0.013	0.000	0.000

Table 26 shows that programs that had a decrease in the grade are positively associated with having faculty with low EC in the period 2010-2012 and negatively associated with programs having faculty with a high EC in the last two periods of evaluation. Moreover, programs which maintained their grade had a negative association with having faculty with high EC in the period 2007-2009, a positive association with having faculty with low EC in the period 2010-2012 and a positive association with having faculty with high EC in the period 2013-2016. Programs that had an increase in its grade had a positive association with having faculty with high EC in the first two periods of evaluation, but in the last period this association was negative. These results show a change of the evolution of the programs in the last period in what regards the EC of the faculty.

Table 26 – Performance vs. EC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.14	-0.23	**-4.01**	**2.59**	1.42	**-2.60**	1.84	0.76
No change	**-2.34**	1.78	0.57	-0.94	**2.42**	-1.48	**4.89**	-1.89	**-3.02**
Grade rose	**2.73**	**-2.23**	-0.50	**4.60**	**-4.79**	0.20	**-3.16**	0.52	**2.65**

Table 27 shows that in all periods a grade decrease is positively associated with having faculty with low WEC. Moreover, a grade increase is negatively associated with having faculty with low WEC in all periods of evaluation.

Table 27 – Performance vs. WEC - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	**2.50**	-1.59	**-4.01**	**3.92**	0.09	**-4.38**	**5.40**	-1.02
No change	-1.37	1.78	-0.40	-0.46	1.46	-1.00	**3.09**	**-2.11**	-0.98
Grade rose	1.73	**-2.73**	0.99	**4.11**	**-5.03**	0.92	0.28	**-2.18**	1.91

Although faculty eigenvector centrality influences program performance, this metric does not provide a clear strategy for how programs can improve performance. In the unweighted case, for instance, the concept increase occurred in programs whose faculty members were highly central

in the first two periods. However, in the last period, the increase in the concept is associated with a shortage of faculty members with high eigenvector centrality.

In the weighted case, it can be stated that to improve performance the programs should not have significant amounts of faculty members with low eigenvector centrality. In order to increase eigenvector centrality, collaboration with the most central faculty in the network should be encouraged.

5.7 PageRank

For the analysis of the PageRank, a dumping factor of 0.85 was adopted (a dumping factor of 0.7 generated similar results). In Table 28, the chi-square test indicated an association of PageRank (PR) with the evolution of the programs' grade only in the first two evaluation periods. In Table 29, where the weighted PageRank (WPR) is considered, only in the second period a significant association was found.

Table 28 – Performance vs. PR - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	11.192	14.950	2.839
df	4	4	4
Asymp. Sig. (2-sided)	0.024	0.005	0.585

Table 29 – Performance vs. WPR - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	8.477	20.145	4.175
df	4	4	4
Asymp. Sig. (2-sided)	0.076	0.000	0.383

In Tables 28 and 29, no significant association was found in the last period of evaluation between evolution of the programs grade and PR or WPR of its faculty. In the triennium 2007-2009, only the unweighted PR showed significant association with the evolution of the programs' grade. In that period, it was verified a positive association between no change in the programs' grade and low PR and also between an increase in the programs' grade and high PR. Still in this first period, a negative association was found between increase in the programs' grade and low PR.

In the period 2010-2012, it was verified that a grade increase was positively associated with high PR and high WPR and negatively associated with low PR and low WPR. Moreover, a negative association was observed between a grade decrease and high PR. Finally, maintaining the grade was positively associated with low WPR and negatively associated with high WPR.

Table 30 – Performance vs. PR - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	1.38	-0.55	**-2.42**	0.48	1.95	-1.42	0.95	0.47
No change	-1.86	**2.58**	-0.85	-1.18	1.94	-0.76	-0.08	-0.08	0.16
Grade rose	**2.23**	**-3.14**	1.07	**3.39**	**-2.39**	-1.01	1.26	-0.70	-0.56

Table 31 – Performance vs. WPR - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.91	2.04	-1.27	-1.63	1.27	0.36	-1.71	1.55	0.17
No change	-1.37	1.41	-0.09	**-2.61**	**2.42**	0.20	0.15	-0.30	0.16
Grade rose	1.73	-2.18	0.55	**4.11**	**-3.59**	-0.53	1.26	-0.95	-0.31

Through PageRank the status or prestige of the faculty members are defined by the prestige of its co-authors. Faculty members prestige, defined by this metric, is not a good indicator for improving program performance because it has behaved erratically over the three periods analyzed, as presented above.

5.8 Utility

Tables 32 and 33 show that an association of the evolution of the program's grade and the (weighted) utility (UT) occurs only in the first two periods. Moreover, it can be seen that this association is stronger with the unweighted utility.

Table 32 – Performance vs. UT - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	11.126	13.370	7.734
df	4	4	4
Asymp. Sig. (2-sided)	0.025	0.010	0.102

Table 33 – Performance vs. WUT - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	10.296	10.045	5.658
df	4	4	4
Asymp. Sig. (2-sided)	0.036	0.040	0.226

From the analysis of the residues shown in Tables 34 and 35, it can be verified that in 2007-2009 the programs that remained with the same grade presented a higher number of faculty with low UT and WUT. In that same period, programs that increased their grades had fewer faculty with low UT and WUT. In the following period, it is noticed that a higher number of faculty with high UT and WUT were part of programs that improved their grade. In the last period, no significant association was observed.

Table 34 – Performance vs. UT - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.94	1.82	-0.87	-1.37	-0.13	1.54	-2.11	1.12	1.01
No change	-1.44	**2.26**	-0.82	-1.66	**2.76**	-1.18	-0.31	-0.63	0.95
Grade rose	1.82	**-2.97**	1.16	**2.91**	**-2.65**	-0.22	2.08	-0.24	-1.86

Table 35 – Performance vs. WUT - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	-0.23	1.82	-1.59	-1.63	0.01	1.64	-2.06	1.12	0.95
No change	-1.13	**2.02**	-0.89	-1.42	1.66	-0.25	0.03	-0.41	0.37
Grade rose	1.24	**-2.73**	1.49	**2.91**	-1.68	-1.24	1.67	-0.48	-1.20

In analyzing the benefit or utility of faculty members belonging to the network structure studied, we find that only for the first two periods, that the absence of faculty members with low utility or the presence of faculty members with high utility can lead to better program performance. To increase a faculty utility, collaboration with coauthors who are dedicated to fewer other coauthors should be encouraged.

5.9 Local Clustering Coefficient

Analyzing the relationship between the local clustering coefficient and the evolution of the program's grade we verified, by chi-square test Tables 36 and 37, that the dependence occurred in the first and second triennial for both networks, unweighted and weighted. Being more pronounced in the unweighted case.

Table 36 – Performance vs. CT - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	11.141	12.432	6.163
df	4	4	4
Asymp. Sig. (2-sided)	0.025	0.014	0.187

Table 37 – Performance vs. WCT - Chi-Square Tests

	Period		
Pearson Chi-Square	2007-2009	2010-2012	2013-2016
Value	9.670	10.588	7.876
df	4	4	4
Asymp. Sig. (2-sided)	0.046	0.032	0.096

In the first period (2007-2009), the analysis of the residuals displayed in Tables 38 and 39 shows that the programs that had an increase in their grade had a higher number of faculty members with medium local clustering coefficient. In the second period (2010-2012), we observed that programs that had a decrease in their grade had a higher number of faculty members with low coefficient of local clustering.

Table 38 – Performance vs. CT - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	0.4	1.1	-1.6	-0.3	**2.3**	**-2.0**	-0.4	1.2	-0.8
No change	1.6	0.8	**-2.4**	1.2	0.0	-1.2	1.7	-2,3	0.6
Grade rose	-1.8	-1,2	**2.1**	**3.1**	**-2.65**	-0.22	-1.5	1.5	0.0

Table 39 – Performance vs. WCT - Adjusted residual analysis

	2007-2009			2010-2012			2013-2016		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
Grade fell	0.4	1.1	-1.6	-1.4	**2.9**	-1.5	-1.1	2.4	-1.3
No change	1.4	0.8	**-2.2**	0.3	0.0	-0.3	0.8	-2.3	1.5
Grade rose	-1.6	-1.2	**2.8**	1.0	**-2.6**	1.6	0.0	0.5	-0.6

The clustering coefficient of a particular author indicates how his collaborators are working together, this measure, however, has not provided any information that helps programs to improve their grades.

5.10 Summary

In order to visualize the effect of the metrics on the performance of the programs by the analysis of the standardized residues, we elaborate graphs for each period, which are displayed in Figures 1 to 6.

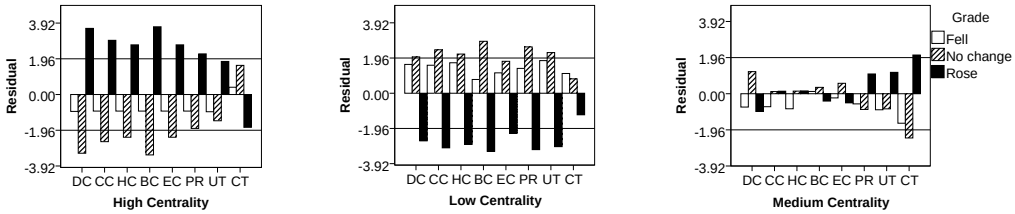

Figure 1 – Standardized residues for unweighted SNA metrics for programs according to their grade evolution in the period 2007-2009.

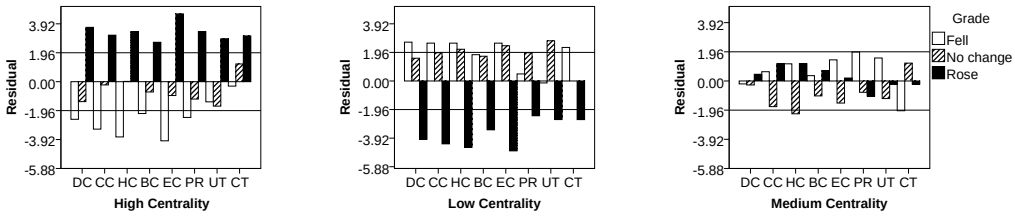

Figure 2 – Standardized residues for unweighted SNA metrics for programs according to their grade evolution in the period 2010-2012.

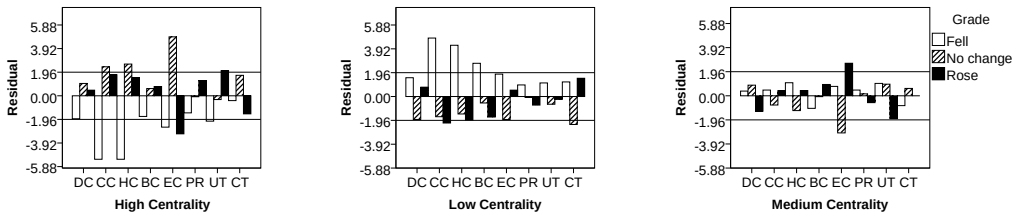

Figure 3 – Standardized residues for unweighted SNA metrics for programs according to their grade evolution in the period 2013-2016.

In the first two evaluation periods, the programs that increased the grade showed a higher number of faculty members with high centrality measures and a lower number of faculty members with low centrality measures, for most of the weighted and unweighted SNA metrics. In the third period of evaluation, no clear pattern was observed in what regards the standardized residues.

The programs that dropped the grade showed a lower number of faculty members with high centrality measures and a higher number of faculty members with low centrality measures for most of the unweighted metrics in all three periods of evaluations. For weighted metrics, in the first period of evaluation, the standardized residues of the programs that dropped the grade were within the expected values. However, in the following two periods of evaluations, the behavior is similar to the case of the unweighted metrics previously mentioned.

In what regards the programs that maintained their grades, for most unweighted metrics, in all three periods of evaluations, the standardized residues were within the expected values. How-

ever, for the weighted metrics, we see that in the first period of evaluation these programs had a lower number of faculty members with high centrality, but in the last period of evaluation these programs showed a higher number of faculty members with high centrality measures. These suggests that the evaluation process of the last period of evaluation seemed to be more severe in comparison to the first period.

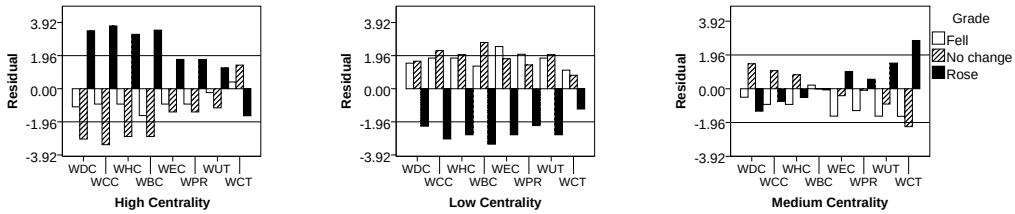

Figure 4 – Standardized residues for weighted SNA metrics for programs according to their grade evolution in the period 2007-2009.

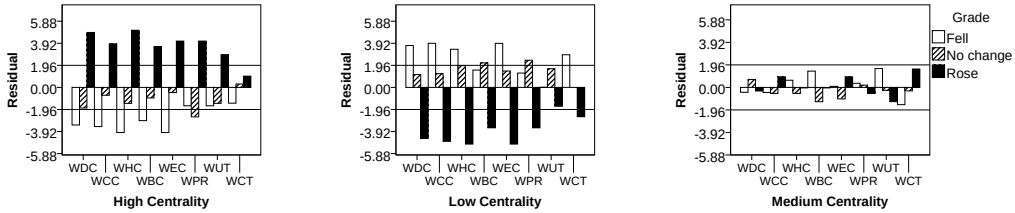

Figure 5 – Standardized residues for weighted SNA metrics for programs according to their grade evolution in the period 2010-2012.

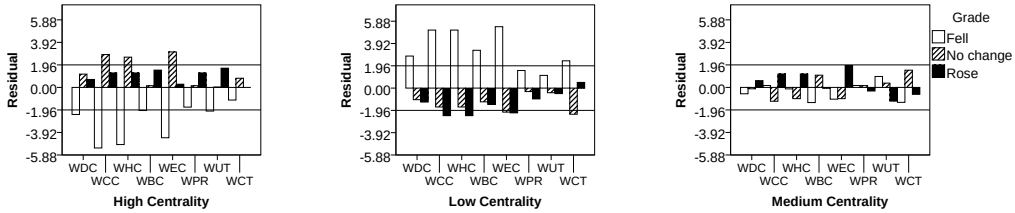

Figure 6 – Standardized residues for weighted SNA metrics for programs according to their grade evolution in the period 2013-2016.

To sum up, in most cases it was shown a positive association with higher SNA metrics and program increase in grade and programs with lower SNA metrics and program decrease in grade. As emphasized in the previous sections, to obtain higher centrality measures programs should encourage their faculty to increase their collaboration with existing coauthors, increase academic collaboration, specially with those researchers further away in the network, increase collaboration with highly central nodes and also to establish collaboration with researchers who have fewer connections in the network.

6 CONCLUSIONS

In order to ensure and maintain the quality of the master's and doctoral programs in Brazil, an evaluation is carried out by the National System of Graduate Programs through ad-hoc consultants and with the participation of the academic community. The evaluation, which took place every three years since 1998, is now carried out every four years since 2015. The system assigns a grade ranging from 1 to 7 following five criteria. The intellectual production is the criterion with greater weight in the definition of the grade. The intellectual production of the program is mainly measured by the journal publications of the permanent faculty of the programs both in terms of quantity and quality.

This work intended to study the co-authorship network of the permanent faculty of graduate programs in Production Engineering in Brazil since 2007. The analysis was divided into three periods corresponding to the evaluation periods. Besides exposing the major characteristics of such networks, an study with chi-square tests was made to investigate the association of SNA metrics of the permanent faculty members with the grades of the programs they participated. The following conclusions were obtained:

- There exists a relationship between the evolution of the program's grade and the number of faculty members who had collaboration in their intellectual production. Furthermore, this association was more intense when the relations of co-authorship were established among professors of the same graduate program (endogenous relation). Residue analysis showed that programs that had a decrease in their grade had a higher number of faculty members which do not collaborate. On the other hand, the programs that maintained or increased their grade had a higher number of faculty members who established co-authorship relations.

- All SNA metrics, studied in this work, presented relationships with the performance of the programs in at least one of the evaluation periods. In most cases, weighted metrics showed stronger associations with the evolution of the program's grade. Thus, not only the number of co-authors contributes to the evolution of a program's evaluation, but also the frequency of such collaborations.

- All programs, in which there was an excess of faculty with high centrality, for any of the metrics studied, weighted or not, had the grade increased or at least maintained. The only exception was for the EC in the period 2013-2016, where a shortage of faculty members with high EC was observed in programs that increased the grade. Nonetheless, faculty members with low centrality were more abundant in programs that had a decrease in their grade.

- To establish a ranking of the metrics regarding the value of the Pearson chi-square test for those who showed dependence on the evolution of the program's grade, we highlight the five main ones in each period. In the first period, they were: WCC, WBC, BC, DC and WHC. In the second period, they were: WHC, WEC, EC, WDC and WCC. Finally, in the

third period, they were: WCC, WEC, WHC, CC and HC. We can observe the predominance of weighted metrics in this ranking. Moreover, WCC and WHC appeared in the rankings in all three periods.

Therefore, through the above-mentioned conclusions, we can affirm that the establishment of co-authoring among faculty, as well as the centrality of the faculty in the co-authoring network, are useful indicators when assessing graduate programs. Other alternatives and useful indicators were obtained in the work of Silva et al. (2017), when analyzing Brazilian Computer Science graduate programs. They concluded that the top-ranked graduate programs were formed by more experienced faculty members, who have mentored more Ph.D. students and who tend to publish more in high-quality venues. The co-authorship network analyzed in this study was formed by permanent faculty members of the Production Engineering Graduate Programs and their collaborators. However, other types of co-authorship networks could be constituted for this analysis, the network could be built only by permanent faculty members, for example. In this way, the metrics would take only the endogenous relations. Alternatively, we could obtain the metrics for the exogenous relations. A future work comparing the relations of these two approaches with the performance of the programs could generate more insights.

In this study, we do not distinguish who are the permanent faculty collaborators. For future work, we could categorize collaborators according, for example, whether the collaborator is a native or a foreigner, a student or another researcher. Thus, one could verify if collaboration with different categories would cause different impacts on the performance of the programs.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001. The first author would like to acknowledge the financial support of the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) and the second author would like to acknowledge the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

[1] ABBASI A, ALTMANN J & HOSSAIN L. 2011. Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, **5**(4): 594–607.

[2] ABBASI A, HOSSAIN L & LEYDESDORFF L. 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, **6**(3): 403–412.

[3] ABRAMO G, CICERO T & D'ANGELO C. 2013. The impact of unproductive and top researchers on overall university research performance. *Journal of Informetrics*, **7**(1): 166–175.

[4] AMORIM R. 2014. *Uma análise das redes de colaboração entre bolsistas de produtividade e entre programas de doutorado em estatística no brasil*. Master's thesis. Universidade Federal de Pernambuco.

[5] ANASTASIOS T, SGOUROPOULOU C, PAPAGEORGIOU E, TERRAZ O & MIAOULIS G. 2012. Co-authorship networks in academic research communities: the role of network strength. pp. 150–155. 16th Panhellenic Conference on Informatics.

[6] ANDRADE R. 2016. *A influência das redes de coautoria na performance dos bolsistas de produtividade e nos programas de pós graduação em engenharia de produção*. Master's thesis. Universidade Federal de Pernambuco.

[7] ANDRADE R & RÊGO L. 2015. Conhecendo a rede de coautoria dos bolsistas de produtividade em pesquisa da área de engenharia de produção ea sua influência no nível de produtividade. In: *XLVII Simpósio Brasileiro de Pesquisa Operacional-SBPO*.

[8] ANDRADE RLD & RÊGO LC. 2017. Exploring the co-authorship network among cnpq's productivity fellows in the area of industrial engineering. *Pesquisa Operacional*, **37**(2): 277–310.

[9] BARABÁSI AEA. 2015. *Network science*. Cambridge university press.

[10] BARRAT A, BARTHELEMY M, PASTOR-SATORRAS R & VESPIGNANI A. 2014. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(11): 3747–3752.

[11] BONACICH P. 1987. Power and centrality: A family of measures. *American Journal of Sociology*, **92**(5): 1170–1182.

[12] BRIN S & PAGE L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, **30**(1): 107–117.

[13] CAINELLI G, MAGGIONI MA, UBERTI TE & DE FELICE A. 2015. The strength of strong ties: How co-authorship affect productivity of academic economists? *Scientometrics*, **102**(1): 673–699.

[14] CAPES. 2014. Caracterização do sistema de avaliação da pós-graduação. Available at: http://www.capes.gov.br/avaliacao/resultado-da-avaliacao-de-programas.html. January 16, 2019.

[15] DE MELLO C, CRUBELLATE J & ROSSONI L. 2010. Dinâmica de relacionamento e prováveis respostas estratégicas de programas brasileiros de pós-graduação em

administração à avaliação da capes: proposições institucionais a partir da análise de redes de co-autorias. Revista de Administração. *Contemporânea*, **14**(3): 434–457.

[16] DIGIAMPIETRI L, MENA-CHALCO J, DE MELO P, MALHEIRO A, MEIRA D, FRANCO L & OLIVEIRA L. 2014. Brax-ray: an x-ray of the brazilian computer science graduate programs. *Plos One*, **9**(4): e94541.

[17] EATON J, WARD J, KUMAR A & REINGEN P. 1999. Structural analysis of co-author relationships and author productivity in selected outlets for consumer behavior research. *Journal of Consumer Psychology*, **8**(1): 39–59.

[18] FIELD A. 2013. *Discovering statistics using IBM SPSS statistics*. University of Sussex.

[19] FREEMAN L. 1978. Centrality in social networks conceptual clarification. *Social Networks*, **1**(3): 215–239.

[20] HAGBERG AA, SCHULT DA & SWART PJ. 2008. Exploring network structure, dynamics, and function using networkx. In: VAROQUAUX G, VAUGHT T & MILLMAN J (Eds.), *Proceedings of the 7th Python in Science Conference*. pp. 11–15. Pasadena, CA USA.

[21] HART RL. 2000. Co-authorship in the academic library literature: a survey of attitudes and behaviors. *The Journal of Academic Librarianship*, **26**(5): 339–345.

[22] JACKSON M. 2008. *Social and economic networks, princeton*. Princeton univ.

[23] JACKSON M & WOLINSKY A. 1996. A strategic model of social and economic networks. *Journal of Economic Theory*, **71**(1): 44–74.

[24] LEE S & BOZEMAN B. 2005. The impact of research collaboration on scientific productivity. *Social studies of science*, **35**(5): 673–702.

[25] LIU J, LI Y, RUAN Z, FU G, CHEN X, SADIQ R & DENG Y. 2015. A new method to construct co-author networks. *Physica A: Statistical Mechanics and its Applications*, **419**: 29–39.

[26] LIU X, BOLLEN J, NELSON ML & VAN DE SOMPEL H. 2005. Co-authorship networks in the digital library research community. *Information processing & management*, **41**(6): 1462–1480.

[27] MARTINS CB, MACCARI EA, MARTINS SB & SERRA FA. 2014. The influence of co-authorship networks on the performance of graduate programs in the Brazilian evaluation system. In: *Proceedings of PICMET'14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration*. pp. 1569–1584. IEEE.

[28] MENA-CHALCO JP & JUNIOR RMC. 2009. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, **15**(4): 31–39.

[29] NEWMAN ME. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences*, **101**(suppl 1): 5200–5205.

[30] ONEL S, ZEID A & KAMARTHI S. 2011. The structure and analysis of nanotechnology co-author and citation networks. *Scientometrics*, **89**(1): 119–138.

[31] ONNELA JP, SARAMÄKI J, KERTÉSZ J & KASKI K. 2005. Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, **71**(6): 065103.

[32] PAGE L, BRIN S, MOTWANI R & WINOGRAD T. 1999. The PageRank citation ranking: Bringing order to the web. Tech. rep.. Stanford InfoLab.

[33] RÊGO LC & DOS SANTOS AM. 2019. Co-authorship model with link strength. *European Journal of Operational Research*, **272**(2): 587–594.

[34] ROCHAT Y. 2009. Closeness centrality extended to unconnected graphs: The harmonic centrality index. Tech. rep.. ASNA, EPFL-CONF-200525.

[35] SILVA TH, LAENDER AH, DAVIS CA, DA SILVA APC & MORO MM. 2017. A profile analysis of the top Brazilian Computer Science graduate programs. *Scientometrics*, **113**(1): 237–255.

[36] TEAM PC. 2015. Python: A dynamic, open source programming language. Python Software Foundation. Available at: https://www.python.org/.

[37] UDDIN S, HOSSAIN L, ABBASI A & RASMUSSEN K. 2011. Trend and efficiency analysis of co-authorship network. *Scientometrics*, **90**(2): 687–699.

[38] WASSERMAN S, FAUST K ET AL. 1994. *Social network analysis: Methods and applications*. vol. 8. Cambridge university press.