

Using Multilayer Social Networks in an Analysis of Higher Education for Professional Demand

Rodrigo Campos¹, Rodrigo Pereira dos Santos² and Jonice Oliveira¹

¹ Federal University of Rio de Janeiro, Rio de Janeiro, 21941-901, Brazil

² Federal University of the State Rio de Janeiro, Rio de Janeiro, 22290-240, Brazil
rodrigo.campos@ufrj.edu.br, rps@uniriotec.br, jonice@dcc.ufrj.br

Abstract. Authorization to open undergraduate courses in major cities considers the quality of physical facilities, faculty, and organization of institutions. An important factor for the process of opening courses is the region economic factor. Considering the regional labor market for such planning can bring benefits and interconnecting these elements transforms cities into smart cities. However, although there are several big data sources that provide this information, there is still an individualistic data view. Therefore, this work proposes to interconnect these factors with the multilayer social networks resources to support the decisions of higher education and their relations with the professional demands. To do so, an experiment is carried out to relate data from higher education offerings and employment/unemployment rates, creating a multilayer graph from these unstructured data. Our contribution is the investigation on how non-structured data can be analyzed in a multilayer perspective for this domain and how to assign proportional weights to the nodes in order to generate weighted graphs.

Keywords: Multilayer Social Networks, Weighted Graphs, Social Network Analysis.

1 Introduction

Smart city is considered whose economy and governance is being driven by innovation, creativity and entrepreneurship, enacted by smart people [1]. Transparency of information enables current governments to build more integrated cities with interconnected data. However, although there are several big data sources that provide this information, there is still an individualistic data view.

In Brazilian education, for example, each university has a council that is responsible for establishing standards. In addition, it can create, expand, modify or extinguish undergraduate courses and graduate programs, i.e., define academic policies [3].

Course creation processes are well planned within the institutions, and in several cases it is guided by internal legislation (when there is complete autonomy) or by governmental decrees. Moreover, this process can be defined by laws that assess aspects such as the importance or necessity of creating a new course and its insertion in institutional, regional and national reality. The planning activity also observes and analyzes

if there are courses of related areas in the own institution and carries out an evaluation of the market and potentials for professional formation [4].

Brazilian Ministry of Labor carries out incremental, comparative research to understand market as well as the economic situation evolution in the educational context – but also for decision-making in other areas. Besides the Brazilian Ministry of Labor, the Brazilian Institute of Geography and Statistics (IBGE) also has research dedicated to producing information about the population insertion in industry. This information is very relevant to both ministries since they can follow the situation of a certain economy sector and also identify situations and numbers related to child labor, migration, fecundity, and other aspects according to information needs. While some professionals who performed trainings migrate to the market, another movement is the professionals exit. Despite that, data provided by several agencies are unstructured data.

One of the concepts strongly used in behavioral sciences in recent years is Social Networks Analysis (SNA). A social network can be characterized as a set of people who interact themselves and are grouped into a representation layer. It was commonly represented in a single layer. With the emergence of new scenarios, it was necessary to represent these interactions in more layers or groups.

For representation proposes, the Multilayer Social Networks can have several connectivity channels. Such networks aim to represent and describe the interconnected systems, where each channel is a layer and the nodes of each layer may have different interactions [2]. Boccaletti et al. [2] state that several types of different actors' relations should be considered in social networks, such as: friendship, neighborhood, kinship, same cultural society, partnership or co-worker etc.

In this context, this work investigates how multilayer graphs can use resources to relate such data structurally and visually to facilitate their analysis. To do so, an experiment is carried out to relate data from higher education offerings and employment/unemployment rates, creating a multilayer graph from these unstructured data. Our contribution is the investigation on how non-structured data can be analyzed in a multilayer perspective for this domain and how to assign proportional weights to the nodes in order to generate weighted graphs. As such, this study can help to understand the reasons for professional's exit based on qualitative analysis of the generated graph.

It is observed that the creation of courses is directly related to the market analysis process, which involves studying the professional's inflows and outflows. However, multilayer social network approaches are little explored in this scenario. Therefore, this work uses multilayer social networks to reveal the impact that the undergraduate courses network has on the economic layer or vice versa, i.e., when the local economy influences decision-making within the higher education institution (HEI). So, we used Multilayer Social Networks analysis to suggest better information visualization, applying these concepts in four important layers of the professional's entry process in the market to establish a better understanding of the technology and the underlying market.

The remainder of this paper is structured as follows: Section 2 presents the research scenario, that is, where this work will be applied. Section 3 discuss the structure and execution of the search method scenario and provides information on the civil construction and plotting steps of the information selected in the methodology. Section 4 dis-

cusses the results found in the experimentation process, validating the previous sections, and interprets the results of some social network analysis metrics. Then, Section 5 describes related work regarding the application of multilayer social networks in a similar scenario. Section 6 concludes the paper and points out future work.

2 Research Scenario: State of Rio de Janeiro

As a study scenario, the city of Rio de Janeiro was selected because, despite being one of the smallest federative units in Brazil, it is the 3rd most populous [5], providing a mass of data of great importance. It is reinforced by government initiatives to turn some cities in the State of Rio de Janeiro intelligent, but they are failed approaches, since these initiatives exempt themselves from improving urban planning, providing data by the public entity, making emergency decision over the preventive [6].

One of these actions that need improvement in its function in the cities, is the *Centro de Operações da Prefeitura do Rio* (Operational Center of the Municipality of Rio de Janeiro), which brought together the data flow of thirty public bodies with emphasis on meteorology, traffic, social media information, processing, visualizing, analyzing and monitoring a large amount of live service data and bringing this information together with reports and public data that are enabled by institutional surveys of government agencies, investing aspects of the city and its growth in an integrated way [1].

These governmental agencies are of great importance not only for surveying aspects of public mobility, but also in the educational and economic context that are emphasized in this work. In the educational area, one can cite as an example the INEP (Brazilian Institute for Educational Research) that presents reports with purposes such as to show statistics for the levels and modalities of higher education, contemplating aspects referring to the conditions of supply, access and participation, efficiency and income and sociodemographic context. INEP databases can be accessed via the Internet, which also contains several information and reports that can be consulted or copied [7].

Other reports are data on the formal work links organized by the Brazilian Ministry of Labor (MT, in Portuguese), and available in the General Register of Employed and Unemployed Persons (CAGED, in Portuguese) and in the Annual Social Information Report (RAIS, in Portuguese), which are the main sources of data capable of describing a recent economic geography of the formal labor market, little known and largely neglected by the mainstream media [8].

This demonstrates a series of very-large bases that are little explored in a unified way. The next sections deal with more detail on the use of these reports and others in the integrated analysis, describing the scenario in the State of Rio de Janeiro, but being sufficient to expand in a national analysis with other states.

3 Research Methodology

This section addresses the experimental work structure. To do so, in the first moment, a question is defined to be answered, delimiting the scope of the work analysis. After-

wards, the main data sources of the Brazilian government that serve to create the multilayer graph are discussed. The reason for choosing each graph is also explained. Then, we discuss how to assign weight to the graph nodes and finally the step of importing this manipulated data to the platform chosen to generate multilayer data.

3.1 Social Network Modeling

Aiming for a quantitative analysis of the indicators in the economic segment of the State of Rio de Janeiro (experimental validation), this article also deals with how the multilayer social networks approach can be explored in a way that helps to better visualize and analyze interconnected data. This work also inserts a qualitative approach, focusing on the advantages and disadvantages of the tool chosen in the experiment. Through this tool and the imported and analyzed data, this research may answer the main research question: *What are the characteristics that can help to analyze relationship between the training of professionals in one market segment from undergraduate courses and economic activity in the region?*

To answer this question, a multilayer graph is created in the next subsections, with four layers: knowledge area, higher education institutions, unemployment, and employment. As the experiment scope, it was defined that data would be filtered by “State of Rio de Janeiro”, in reason of the advantages and motivations described in Section 2. Moreover, data would be filtered by “not all courses”, but only the courses of the civil construction sector. It is just to reduce the scope of study, optimizing the data input time in the multilayer network analysis platform. As the proposal is the application of the multilayer network approach to support data interconnection, any segment could be selected, with no difference in the ultimate goal. In addition, data were collected from the year of 2014. This allows a more accurate analysis and ensures that this experiment can be conducted with other variables.

Graphs under analysis of social networks can be visualized by creating a relationship between two or more individuals, where each individual is called a node. These relationships are represented as edges, which in turn can be directed or not directed [9]. In this case, the graph layers were designed to be:

Layer I) Knowledge Area: is a clique, meaning all nodes are linked to all nodes. Each node represents a knowledge (course) area. Each node also has connection to layer II, indicating the HEIs that have course in an area (red edge in Figure 1(d)). In our experiment, only one node has this link, which represents the node of the civil construction knowledge area.

Layer II) Higher Education Institutions (HEI): each node is a higher education institution in the State of Rio de Janeiro. As can be seen in Figure 1(c), nodes are linked to higher education institutions that have at least one course in the civil construction area that is in common. Each node connects to the layer III nodes (municipalities) where the institution has a course (red edge in Figure 1(c)).

Layer III) Unemployment: each node is a municipality (Figure 1(b)) of the State of Rio de Janeiro (total of 92). Nodes are linked in the graph to all neighboring municipalities. This graph is the same of layer IV (employment), so the difference is on the weights as described next, and the connection between these layers is optional.

Layer IV) Employment: as can be seen in the Figure 1(a), each node is a municipality of the State of Rio de Janeiro (this State has a total of 92). Nodes are linked in the graph to all neighboring municipalities. It is a graph representing the geography of the State.

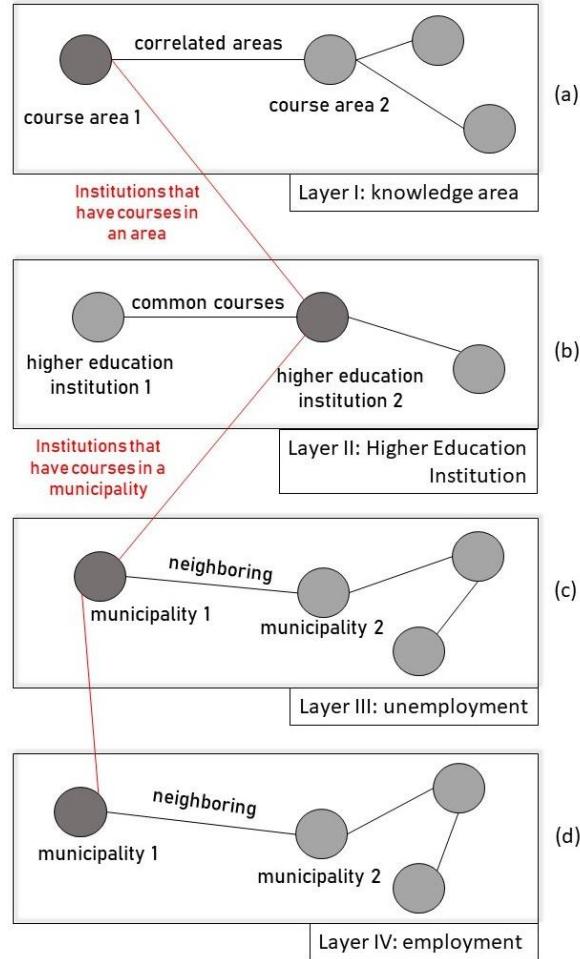


Fig. 1. Concept of layer structure, nodes and links

The last three layers are weighted. The weight at the edges gives the graph a representation of, for example, how much a relation between two (if vertices represent people, the edge weights can vary from higher to familiar people and lower to lesser known), or distances (if the vertices represent cities, the edge weights can represent the distance between them or the travel cost) [10].

Edges have weights. To define such weights in layer IV and III in a unified and reproducible form, a very important economic indicator called the Locational Quotient was used [11]. In the layer IV case (employment by municipality), it is expected to analyze employability fees, then the variables for calculating the Locational Quotient

were the number of employees in the civil construction sector by municipality in 2014. The layer III exists with the same logic as the last layer. However, for the edge's weight, the variable introduced in the Locational Quotient method was the number of unemployed professionals in the civil construction sector per municipality in 2014. For the higher education institutions graph (layer II), the variables were the number of courses.

As such, the layers represent: i) knowledge areas ii) higher education institutions grouped by the number of courses in the selected area iii) unemployment statistics in municipalities, and iv) the market network grouped by municipalities .

As described in this section, layers IV and III can be represented with the same graph, since nodes are linked in the two graphs to all neighboring municipalities and difference is on the weights. The next subsections discuss, therefore, how the process of construction of these graphs in the perspective of multilayers is.

3.2 Analysis Tool Selection

Normally, multilayer networks encode information using individual layers separately. However, to satisfy operational requirements and to be able to apply an effective analysis of complex systems, it is of the utmost importance to also reproduce these results, preferably in a specific open source software to visualize multilayer networks. Thus, it can represent the analysis results of these networks significantly.

Several tools are currently found as multilayer social networks facilitators. For example, Pymnet is a multilayer networks library for Python that has several network analysis methods, transformations, read and write networks, and a scalable implementation for sparse networks, that is: memory usage is staggered linearly with the number of edges and the number of nodes. In addition, it has a view using Matplotlib or D3 [12]. Another tool example is MAMMUL, with an approach to metrics and models for multilayer networks using C.

However, this research identified a tool with application for R language that could bring all the expected resources: MuxViz [13]. This choice was also motivated by the fact that MuxViz is free, open source software that allows a more interactive visualization and a graphical interface with more room for greater network control. MuxViz is also based on the GNU Octave language and can run on Windows, Linux or Mac OS X systems, which ensures greater portability.

3.3 Data Collect

For better applicability of the validation experiments, data previously collected by several Federal and State agencies in the State of Rio de Janeiro were analyzed. We noticed information from all administrative regions, especially when we aggregated the entire data mass for a more objective view.

As shown in Table 1, more than one data source were found that are able to fill the layer IV. However, after an analysis of these sources content, the one chosen to represent the first layer was RAIS (Annual Relation of Social Information) microdata, belonging to the Brazilian Ministry of Labor and instituted by Decree No. 76.900 of 12/23/75. This relationship purpose is to control labor activity in Brazil, as well as to identify the Brazilian workers who are entitled to the salary bonus. However, RAIS is

also very useful for the labor statistics elaboration, since data are collected throughout the national territory based on the private and public domain (direct or indirect, Federal, State or Municipality administration), as well as temporary workers, non-effective servants, union leaders and other categories established in the decree [14].

For analysis purposes, data from the State of Rio de Janeiro were extracted from RAIS and structured by municipalities. The municipalities list of the State of Rio de Janeiro was extracted from the Municipalities Codes table [15] elaborated by the IBGE (last updated in 2015). This table is not only used by the IBGE, but also by other institutions, such as the Brazilian Ministry of Labor, thus being a national standard.

Table 1. Selected data source¹

| Name | Government Department | Source | Data availability period | Layer |
|---|-----------------------|---|--------------------------|-------|
| <i>InepData – Educação Superior</i> | INEP | http://portal.inep.gov.br/inep-data | 2009 to 2014 | I/II |
| <i>Boletim de Conjuntura Econômica</i> | CEPERJ | http://www.ceperj.rj.gov.br/Bolcon/bolcon.html | Jan. 2009 to Feb. 2017 | IV |
| <i>Pesquisa Nacional Por Amostra de Domicílios Contínua</i> | IBGE | http://www.ibge.gov.br/home/estatistica/indicadores/trabalhoerendimento/pnad_continua/default.shtml | 2012 to 2017 | IV |
| <i>Relação Anual de Informações Sociais</i> | RAIS | http://pdet.mte.gov.br/rais/rais-anos-anteriores | 2007 to 2015 | IV |
| <i>Cadastro Geral de Empregados e Desempregados (CAGED)</i> | MT | http://portafat.mte.gov.br/programas-e-acoes-2/caged-3/dados-estaduais/ and http://bi.mte.gov.br/eec | 2009 to 2017 | III |

The Brazilian Ministry of Labor through RAIS makes employment data available annually. In addition to RAIS, the Brazilian Ministry of Labor also provides the General Register of Employed and Unemployed (CAGED, as showed in Table 1). It is used as a permanent employee's dismissal record and serves as a basis for the preparation of studies, research and projects of the Brazilian Ministry of Labor and other bodies linked to the labor market. Therefore, data from both bases were used to perform the employability and unemployment survey.

Next, data were extracted from the Brazilian Ministry of Labor (MT) portal and imported into a relational database. Once defined, the base had the following tables: municipality, neighboring, employment, LQ_neighboring_employment, HEI, HEI_municipality, course_area, unemployment, and LQ_neighboring_unemployment.

¹ Names are presented in the original language, in Portuguese. It can be translated (in order of display) to: i) InepData - Higher Education, ii) Economic Bulletin, iii) National Household Sample Survey Continues, iv) Annual Social Information Report, and v) General Register of Employed and Unemployed Workers (CAGED)

3.4 Calculating the Edge Weight

According to Scherer and Moraes (2012), the Locational Quotient (LQ) indicates the relative concentration of a certain branch of activity "i" in a region "j", compared to the participation of the same branch in the State. As such, the higher LQ, the greater the specialization of the region in the respective activity branch. LQ can be analyzed from specific branches, or as a whole.

The last three layers are weighted graphs, that is, their edges have weights. Therefore, LQ suggested by Scherer and Moraes (2012) was used in layer IV and III to find a real number that measures how much that connection weighs. With this objective established, LQ of each municipality was calculated for the last layer, which deals with the employability among the municipalities. Taking the total number of employees (EMP) of each municipality informed by RAIS as a variable, LQ calculation of employability by municipality (LQ_EMP_MUN) was performed as follow:

$$\text{LQ_EMP_MUN} = (\text{EMP of sector } i \text{ in the municipality } j / \text{EMP in the municipality } j) / (\text{EMP of sector } i \text{ in the country} / \text{EMP in the country}) \quad (1)$$

Since the research scope was defined for the year 2014 and knowing that sector $i =$ civil construction, the total EMP Country sector i is 2,815,686 and total EMP in the country is 48,948,433. From these values together with the number of employees in the sector of the municipality and the total number of employees in this municipality, LQ_EMP_MUN was obtained. The second process stage was performed in two nodes union. The nodes (municipalities) were joined when they were neighboring municipalities, then the weight of this connection was defined as the sum of LQ_EMP_MUN of the two connected municipalities, representing the strength of employability between those two municipalities.

The same process was used for the layer III. The formula had the same structure, but the input values represented the number of unemployed (UNEMP) in the civil construction sector, in the municipalities of the State of Rio de Janeiro in 2014. The formula can be defined as:

$$\text{LQ_UNEMP_MUN} = (\text{UNEMP of sector } i \text{ in the municipality } j / \text{UNEMP in the municipality } j) / (\text{UNEMP of sector } i \text{ in the country} / \text{UNEMP in the country}) \quad (2)$$

From the values found for each municipality, LQ_UNEMP_MUN of each municipality was calculated, considering that the total UNEMP Country sector i is 2,838,611 and total UNEMP in the country is 21,368,062. Again, the edges weight is calculated by the sum of LQ_UNEMP_MUN of the two nodes (municipalities) related in the layer.

The layer II formula consisted in using a different logic. Considering that each edge connects two HEI, the formula sums the number of courses in area i of these HEI and verifies the proportion that this sum represents on the total of courses in the area in the selected region (Rio de Janeiro), thus returning a percentage or edge's weight (WEIGHT_COURSE_HEI). This means that the applied formula was:

$$\text{WEIGHT_COURSE_HEI} = (\text{sum of courses of sector } i \text{ in the HEIs} / \text{courses of sector } i \text{ in the selected region}) * 100 \quad (3)$$

Given the formula, it was identified that, from 4,543 courses related to civil construction area in Brazil, 348 were in Rio de Janeiro (courses of sector i in the selected region) in 2014. Therefore, for each edge, the formula application was performed considering the edges' two nodes.

3.5 Layers Preview

With the layers, nodes, relationships and defined edges' weights, we reached the stage of importing this information into the visualization and analysis platform chosen in this research (MuxViz) [13]. Besides being free, the tool has a forum with several feedbacks from users regarding its experience of use, ensuring useful information from the installation process to the use of more advanced level, and error correction.

Therefore, the first step was the experimental environment preparation. To do so, the following elements were installed: i) Ubuntu Linux Virtual Machine 16.04LTS, ii) GNU Octave in version 3.4, iii) R in version 3.2.3, iv) MuxViz, and v) Geospatial Data Abstraction Library (GDAL). The installation process could also be performed on the Windows system. However, more failure chances were identified in this system, then we preferred to use a Linux distribution.

The following is the evaluation carried out to analyze the efficiency and behavior of the proposed solution, besides performing the validation.

4 Results and Evaluation

In order to show Locational Quotient effectiveness as weight to multilayer graphs and of the not structured data analysis in the multilayer perspective, the application of real data used to create the four layers graph was very important for the work validation. As such, this step seeks to evaluate if the solution can achieve the objective of making unstructured data analyzable in the multilayer perspective for the defined domains and also the attribution of proportional weights to the nodes to generate weighted graphs.

92 municipalities and 52 different HEIs with courses from the civil construction area (one of the 8 areas) were imported. As each municipality, HEI, and knowledge area represent a node, the four layers of this network have 152 distinct nodes that are related through 1057 edges with different weights. Layers IV and III each have 220 edges. It means that there are 220 frontiers between municipalities in the State of Rio de Janeiro.

Table 2. LQ calculation of some layer IV nodes

| Municipality (j) | EMP sector i in j | EMP in j | LQ of j (applying Equation 1) |
|--------------------------|-------------------|-----------|-------------------------------|
| Duque_de_Caxias | 7,441 | 172,977 | 0.747821 |
| Rio_de_Janeiro_(capital) | 166,362 | 2,654,076 | 1.089671 |
| Magé | 3,481 | 24,500 | 2.469975 |
| Miguel_Pereira | 254 | 4,322 | 1.021653 |
| Nova_Iguacu | 3,637 | 105,180 | 0.601125 |
| Petrópolis | 3,716 | 75,223 | 0.858775 |
| São_João_de_Meriti | 1,096 | 56,836 | 0.335229 |

Table 3. Edge weight calculation of some layer IV edges

| Municipality (k) | Municipality (l) | LQ of k | LQ of l | Weight Edge (sum LQ) |
|------------------|--------------------------|---------|----------|-------------------------|
| Duque_de_Caxias | Rio_de_Janeiro_(capital) | 0.74782 | 1.089671 | 1.837492 |
| Duque_de_Caxias | Magé | 0.74782 | 2.469975 | 3.217795 |
| Duque_de_Caxias | Miguel_Pereira | 0.74782 | 1.021653 | 1.769474 |
| Duque_de_Caxias | Nova_Iguacu | 0.74782 | 0.601125 | 1.348946 |
| Duque_de_Caxias | Petrópolis | 0.74782 | 0.858775 | 1.606596 |
| Duque_de_Caxias | São_João_de_Meriti | 0.74782 | 0.335229 | 1.083050 |

Table 2 has an example with some layer IV nodes. The respective edges and weights are presented in Table 3. Some layer III nodes are exemplified in Table 4 and their edges are shown in Table 5. For layer II, Table 6 and Table 7 exemplify nodes and edges, respectively. Table 6 shows also that one of the records had no reported value (NRV). This occurred with some municipalities that had almost no courses, employees or unemployed. All examples select relations of Duque de Caxias municipality. The same process was carried out with the other 91 municipalities.

Table 4. LQ calculation of some layer III nodes

| Municipality (j) | UNEMP sector i in j | UNEMP in j | LQ of j (apply-ing Equation 2) |
|--------------------------|---------------------|------------|--------------------------------|
| Duque_de_Caxias | 10,047 | 82,866 | 0.912681569 |
| Rio_de_Janeiro_(capital) | 124,911 | 1,039,803 | 0.904292383 |
| Magé | 1,013 | 7,826 | 0.974381109 |
| Miguel_Pereira | NRV | NRV | 0 |
| Nova_Iguacu | 3,671 | 42,902 | 0.644118985 |
| Petrópolis | 3,559 | 31,097 | 0.861526713 |
| São_João_de_Meriti | 11,119 | 30,425 | 2.751024113 |

Table 5. Edge weight calculation of some layer III edges

| Municipality (k) | Municipality (l) | LQ of k | LQ of l | Weight Edge (sum LQ) |
|------------------|--------------------------|-----------|----------|-------------------------|
| Duque_de_Caxias | Rio_de_Janeiro_(capital) | 0.9126815 | 0.904292 | 1.816974 |
| Duque_de_Caxias | Magé | 0.9126815 | 0.974381 | 1.887063 |
| Duque_de_Caxias | Miguel_Pereira | 0.9126815 | 0 | 0.912681 |
| Duque_de_Caxias | Nova_Iguacu | 0.9126815 | 0.644118 | 1.556801 |
| Duque_de_Caxias | Petrópolis | 0.9126815 | 0.861526 | 1.774208 |
| Duque_de_Caxias | São_João_de_Meriti | 0.9126815 | 2.751024 | 3.663706 |

Table 6. LQ calculation of some layer II nodes

| HEI (j) | Courses sector i in j |
|--|------------------------------|
| Universidade Federal do Rio de Janeiro | 22 |
| Centro Universitário Metodista Bennett | 1 |
| Instituto Federal de Educação, Ciência e Tecnologia Fluminense | 7 |
| Centro Universitário Geraldo di Biase | 8 |
| Instituto Tecnológico e das Ciências Sociais Aplicadas e da Saúde do Centro Educ. N. SR ^a Auxiliadora | 3 |
| Faculdade Redentor | 5 |
| Centro Universitário Fluminense | 1 |

Table 7. Edge weight calculation of some layer II edges

| First HEI (l) | Second HEI (m) | sum of courses of HEI (l and m) | Weight Edge (applying Equation 3) |
|--|--|--|--|
| Universidade Federal do Rio de Janeiro | Centro Universitário Metodista Bennett | 23 | 6.60 |
| Universidade Federal do Rio de Janeiro | Instituto Federal de Educação, Ciência e Tecnologia Fluminense | 29 | 8.33 |
| Universidade Federal do Rio de Janeiro | Centro Universitário Geraldo di Biase | 30 | 8.62 |
| Universidade Federal do Rio de Janeiro | Inst. Tec. e das Ciências Sociais Aplic. e da Saúde do Centro Educ. N. SR ^a Auxiliadora | 25 | 7.18 |
| Universidade Federal do Rio de Janeiro | Faculdade Redentor | 27 | 7.75 |
| Universidade Federal do Rio de Janeiro | Centro Universitário Fluminense | 23 | 6.60 |

4.1 Data Results

Some metrics were analyzed through the MuxViz platform. An analysis was regarding density: “the ratio between vertices and nodes characterizes whether the graph is dense or sparse. Dense graphs have many connections per node and sparse graphs have few” [17]. The employment and unemployment layers density are respectively 0.025 and 0.037. The HEI layer has a density of 0.61, which can be explained by the fact that each HEI have many connections with other HEI.

Other measures can be observed through the analysis tool. For example, the graphs’ diameters are 18.4 (for layers IV and III) and 1.6 (on the layer II). The diameter corresponds to the greatest distance between two vertices in a graph [17]. Considering that the scope of educational institutions for the selected specific courses is smaller, the tendency is that the distance between nodes is smaller.

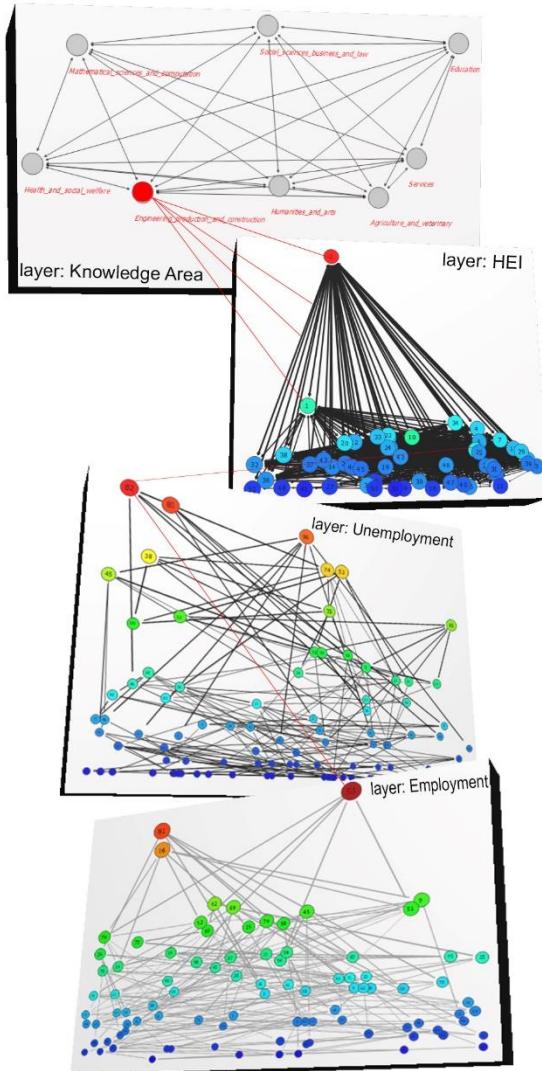


Fig. 2. Plotted layers based on civil construction courses data in 2014
in the State of Rio de Janeiro

As shown in Figure 2, the nodes of layers IV and III are the same, but the structure is different. This occurs because both layers use the same nodes (municipalities). The only difference is its edges' weights, which totally change the value of each link and consequently the format. Figure 2 also shows that layer II has a different format, since only HEIs' courses in the civil construction area are shown (total of 52 HEIs).

4.2 Centrality Results

Although the layers displayed in Muxviz, we can observe the nodes organization according to what was previously exported. The software also offers the diagnostic feature, with countless statistics in relation to the dataset. In civil construction data, it is observed that municipalities such as Itaborá and Itaguaí lead centrality by node degree in layer IV while Seropédica and Itaguaí lead in layer III. The global centrality of a vertex, also known as degree of proximity is the sum of the smallest path between a vertex and the other vertices of the network. A vertex that has the smallest sum of the smallest distances is closer to the others. That is, the greater the global centrality, the greater the distance from one member to the other [9]. Then, this means that these municipalities were the ones that had the most group of employees and unemployed. In layer II the lead centrality by node degree was the Estácio de Sá university.

Further analysis of the selected data can still be performed. Although the number of unemployment was not so high in the municipality of São João da Barra, 64% of the unemployed people were in the area of civil construction. It was observed that this municipality presented the highest proportion of unemployed in the area. This fact indicates that this municipality had a large representation in this segment, and that some factors directly affected unemployment in that area. Observing layer III, the municipality is not connected to any university node, indicating that does not have a course in the civil construction area. However, it is known that it has a border with São Francisco de Itabapoana and Campos dos Goytacazes, which in turn has 8 HEIs with courses in this area to meet the demands of the municipality of São João da Barra.

The municipality with the most courses in the civil construction area is Rio de Janeiro, with 27 institutions that together offered 168 courses in the area. The two universities with the most courses in the area are respectively Estácio de Sá University with 41 courses, and the Federal University of Rio de Janeiro with 20 courses. Meanwhile, the municipality showed 124,911 unemployment and 166,362, that is, a 13.3% increase in the number of jobs.

Meanwhile, in the municipality of Barra Mansa, only 3.8% of the municipality's employees were in the civil construction area. Even so, the University Center of Barra Mansa offered 8 courses in the area in the year 2014. The same happens with Bom Jesus do Itabapoana, which only has 2.94% of employees in the civil construction area and even offered courses in the area.

5 Related Work

Some studies have already investigated the issues addressed in this paper, such as Bloechl et al. [18], which answered which economy sectors are more central to reduce the network complexity by applying "Random walk centrality" and "count-betweenness"-based metrics. Despite applying analyzes based on centrality metrics, this work does not present an approach based on multilayer social networks and only addresses a data context of the economy. In addition, the educational domain was not investigated. On the other hand, the work of Finn et al. [19] presents multilayer approach in weighted

graphs using the same tool presented in our work. However, the domain is totally different as well as the strategies of assigning weight to the nodes accordingly as well.

6 Discussion and Conclusion

After collecting and analyzing the results, we can observe that our work contributes to the use of multilayer social networks in the educational domain, identifying a way of applying weight to the graph nodes. In addition, it defines better ways to apply datasets to multilayer network analysis tools (optimizing these networks) and node connections visualization.

A limitation of our work was to feed the neighboring municipalities data (item 2, section 3.3). This information was made available at random in the IBGE database, thus not providing an XML file with the relation municipality-neighboring municipality. As such, the inserts in the database were performed manually. Another important observation was the step of excluding repeated values. That is, if the first register had the relation Angra_dos_Reis X Mangaratiba and the thirtieth had the relation Mangaratiba X Angra_dos_Reis, only one of these registers could exist.

Besides the example explored in this paper – the civil construction area in the State of Rio de Janeiro –, the multilayer network also allows to obtain a geographic model, facilitating the network visualization according to the exact nodes geolocation. Among the many application possibilities, it was observed that the multilayer social networks analysis becomes more transparent and more understandable when using a specific tool for analyzing this network type (MuxViz). Moreover, MuxViz is an open and free tool.

Based on all the collected information, a future work was identified: to select a greater number of economy sectors to be able to make a comparison among municipalities and to extend inferences. One can also think of implementing the concept of time-varying graphs so that such analysis is not performed only in a specific year, thus generating data more coherent with reality as well as being able to calculate probabilities on that subject.

Acknowledgement

The first author thanks the Federal Institute of Education, Science and Technology of Rio de Janeiro for supporting this research. The second author thanks DPq/UNIRIO for partially supporting this research. Also, the researchers thank CAPES, CNPq and FAPERJ (Brazil) for their financial support to the research group.

References

1. Kitchin, R.: The real-time city? Big data and smart urbanism. *GeoJournal*. 79, 1–14 (2014).
2. Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M.: The structure and dynamics of multilayer

- networks. (2014).
3. Centro Federal de Educação Tecnológica Celso Suckow da Fonseca: Regulamento do Conselho de Ensino, Pesquisa e Extensão (CEPE). (2018).
 4. Stallivieri, L.: O Sistema De Ensino Superior Do Brasil Características, Tendências E Perspectivas | Flacso. (2006).
 5. Instituto Brasileiro de Geografia e Estatística: IBGE Cidades, <https://cidades.ibge.gov.br/brasil/rj/panorama>.
 6. de Souza, A.L.M.: Um Estudo sobre o Conceito de Cidades Inteligentes na Região Metropolitana do Rio de Janeiro, (2017).
 7. Rigotti, J.I.R., Cerqueira, C.A.: As bases de dados do INEP e os indicadores educacionais: conceitos e aplicações. Livros. 71–88 (2015).
 8. Matos, R., Ferreira, R.N.: Brasil em Crise e o Emprego Formal no Sudeste. Caminhos Geogr. 18, 150–164 (2017).
 9. França, T.C., de Faria, F.F., Rangel, F.M., de Farias, C.M., Oliveira, J.: Big Social Data: Princípios sobre Coleta, Tratamento e Análise de Dados Sociais. Tópicos em Gerenciamento Dados e Informações. 8–45 (2014).
 10. Portela, E.: Análise de Dinâmica de Redes Sociais: Aplicação a uma Rede de Preferências Musicais. (2014).
 11. Karla, J., Lima, M., Esperidião, F.: Uma análise dos Quocientes Locacionais das regiões brasileiras nos anos 1991, 2000 e 2010. 18, 175–196 (2014).
 12. Multilayer Networks Library for Python (Pymnet) — Multilayer Networks Library 0.1 documentation, http://www.plexmath.eu/wp-content/uploads/2013/11/multilayer-networks-library_html_documentation/index.html.
 13. De Domenico, M., Porter, M.A., Arenas, A.: MuxViz: a tool for multilayer analysis and visualization of networks. J. Complex Networks. 3, 159–176 (2015).
 14. Ministério do Trabalho: RAIS, <http://www.rais.gov.br/sitio/sobre.jsf>.
 15. IBGE/Concla: Classificações por tema, Tabelas de código de áreas, <http://concla.ibge.gov.br/classificacoes/por-tema/codigo-de-areas/codigo-de-areas.html>.
 16. Scherer, G.J.W., Moraes, S.L.: Análise Locacional das Atividades Dinâmicas do Estado do Rio Grande do Sul. 6º Encontro Econ. Gaúcha. (2012).
 17. Mariani, A.C.: Conceitos Básicos da Teoria de Grafos, <https://www.inf.ufsc.br/grafos/definicoes/definicao.html>.
 18. Bloechl, F., Theis, F.J., Vega-Redondo, F., Fisher, E.O.: Which sectors of a modern economy are most central? CESifo Work. Pap. Ser. 1–13 (2010).
 19. Finn, K.R., Silk, M.J., Porter, M.A., Pinter-Wollman, N.: Novel insights into animal sociality from multilayer networks.