




Development of a relevance index for the subjects in a study plan - Case study: Systems and Computer Engineering, Universidad Tecnológica de Pereira

Desarrollo de un índice de relevancia de las asignaturas de un plan de estudios - Caso de estudio: Ingeniería de Sistemas y Computación, Universidad Tecnológica de Pereira

J. Guerrero-Eraza  ; G. Grandas-Aguirre  ; J. Castaño-Gómez 

Abstract—This document presents the development of an index that aims to quantify, according to some criteria known in graph theory, how relevant a subject is, taking into account its location in the curriculum, its number of credits, its prerequisites and the subjects dependents. The first thing was to model the academic plan using a graph, which considers only two things: the assigned credits and the prerequisites that must be met before taking the subjects. After having this model, graph theory algorithms were applied that allow to measure the importance of a subject with respect to the location in its curricular mesh (Centrality) and allow to give a measure of the importance of the subjects based on academic credits, its prerequisites and subjects depending on it (Neighborhood). It is important to note that the analysis presented is not intended to indicate that one subject is more important than another for the student's professional development, but rather to analyze, in an estimative way, which subjects contribute more to the connectivity of the program and academic flow by this network only taking into account the information found in the curriculum. The result obtained is a composite index, which allows visualizing the relevance degree of the subjects in the study plan.

Index Terms— Educational Data mining, Graph centrality, Graph theory, HITS algorithm, Relevant subjects, Study plan

Resumen— En este documento se presenta el desarrollo de un índice que pretende cuantificar, según algunos criterios conocidos en teoría de grafos, que tanta relevancia tiene una asignatura, teniendo en cuenta su ubicación en la malla curricular, su número de créditos, sus prerequisites y las asignaturas dependientes. Lo primero fue modelar el plan académico mediante un grafo, el cual tiene como elementos considerados únicamente los créditos asignados y los prerequisites que se deben cumplir antes de cursar las asignaturas. Luego de contar con este modelo, se aplicaron algoritmos de teoría de grafos que permiten medir la

importancia de una materia con respecto a la localización en su malla curricular (Centralidad) y permiten dar una medida de la importancia de las asignaturas basada en los créditos académicos, sus prerequisites y asignaturas dependientes de ésta (Vecindad). Es importante rescatar que el análisis presentado no tiene como objeto indicar que una asignatura tiene más importancia que otra para el desarrollo profesional del estudiante, sino por el contrario analizar, de una manera estimativa cuales asignaturas aportan más a la conectividad del programa y al flujo académico por esta red únicamente teniendo en cuenta la información que se encuentra en el curriculum. El resultado que se obtiene es un índice compuesto, que permite visualizar la relevancia de las asignaturas en el plan de estudios.

Palabras claves—Algoritmo HITS, Asignaturas relevantes, Centralidad de grafos, Plan de estudios, teoría de grafos.

I. INTRODUCTION

The curriculum of a college program is made up of subjects, structured in a curricular mesh that makes up a network and has different transit routes. Each subject has a weight represented in academic credits. At the same time, the subjects have prerequisites and are prerequisites for others.

Thus, a subject may trigger different paths or not, it may be a prerequisite for many or few subjects, they have different academic credits, which influences the transit of students through the study plan.

The relevance of the subjects is not simple to visualize in the curricular grid because it is given by the intermediation with other subjects and their academic credits.

This manuscript was submitted on May 27, 2020 and accepted for publication on September 28, 2020.

J. Guerrero-Eraza is Academic Vice-rector of from Universidad Tecnológica de Pereira

G. Grandas-Aguirre is Computer Science Engineering of from Universidad Tecnológica de Pereira.

J. Castaño-Gómez is Computer Science Engineering of from Universidad Tecnológica de Pereira.



The presented article shows how a methodology to address the problem described previously is proposed during an investigation project called *Descriptive model based on decision trees to prioritize subjects in the study plan and find possible academic routes* conducted in “Universidad Tecnológica de Pereira”. During the development of this project, a need arose to rank subjects by relevance in the study plan of each engineering program available at the university.

In response to this need, it was proposed that each program could be represented in a graph, using subjects as nodes and the correlation between them as edges in order to create an index that allows quantifying the relevance of each subject in the study plan. To create the described index, three numerical measures were used, product of two algorithms: HITS algorithm and Betweenness centrality measure. The investigation then took the course of proving the following hypothesis:

“Graph techniques allow determining the relevance of the subjects in the study plan by quantifying the intermediations based on prerequisites and academic credits through the HITS algorithm and Betweenness Centrality measure.”

The proposed methodology was first applied to the Systems and Computer Engineering program at “Universidad Tecnológica de Pereira”, but was then replied and contrasted to all the other engineering programs available in the university. The development of how the process was applied to Systems and Computer Engineering program is described in detail, and then its results are contrasted with the results of the other engineering programs. Nevertheless, it is worth mentioning that because of the nature of how the whole system was applied, it is replicable to any other university program.

After getting the relevance indexes for all the engineering programs, results were validated through statistical analysis using historical records of students about approval or failure in subjects and the number of semesters they took to complete their study plan. This allowed to corroborate the validity of the presented methodology.

II. HYPOTHESIS

Graph techniques allow determining the relevance of the subjects in the study plan by quantifying the intermediations based on prerequisites and academic credits through the HITS algorithm and Betweenness Centrality measure.

III. STATE OF THE ART

S. Saha Ray [1] defines a Graph as two finite sets $G = (V, E)$ where V is the set of vertices of the graph, which is a non-empty set of elements and E the set of edges of the graph which represents the connectivity between two nodes belonging to V . To these two finite sets a third set C can be added, this set will represent the weights existing between an edge.

Another concept that will be useful in this work is the definition of degree of a vertex denoted as $d(v)$ where v is a vertex that belongs to V and represents the number of edges incident with vertex v , this definition is important in of our

context, since it will allow us to know the number of connections existing between the vertices of the graph.

Approaches to the problem were found in the consulted literature in different domains, mainly in the integrity and flow domains in different transport and communication networks.

Guze [2] makes an approach to the problem in transport networks, in his work he creates some definitions that were extrapolated to this problem, these definitions are related to the concept of “Betweenness Centrality” and “Hub Dependence”.

Other similar problems are those that involve analysis of connections between internet sites, Chris Dinga, Xiaofeng Hea, Parry Husbandsa, Hongyuan Zhab and Horst D. Simona [3] use two algorithms to perform this type of analysis and create a composite index to rank the analyzed websites. These algorithms were used because of their similarity to the connections between the various subjects in the curriculum.

For the case study, the proposed modeling consists in representing the subjects that make up the curriculum through vertices, the edges represent the connections between the subjects, and finally, the number of credits existing between two subjects will represent the weight of the edge between them.

IV. METHODOLOGY

The model was structured with a weighted directed graph with the following elements:

- **Graph Vertices (V):** The codes of the subjects of the program.
- **Graph Edges (E):** Indicates that one subject is related to another. In the case of two subjects (A and B) this relationship may represent that A is a prerequisite for B or that B depends on A to be able to take it.
- **Weight of the Edges (C):** It is the number of credits of the prerequisite, example: A and B are two subjects, where A is a prerequisite for B and has a weight of 5 credits, the connection between A and B would be denoted by 5 representing the number of credits that exist between that relationship.

By modeling the problem in this way, an abstraction of the information from the curriculum that is most relevant to the analysis can be done.

A. “Betweenness Centrality”

Betweenness centrality is a way of detecting the amount of influence that a vertex has on the flow of information in a graph [5]. It is often used to find vertices that serve as a bridge from one part of a graph to another.

This centrality of intermeditation, denoted as $cB(v)$ for a vertex v is the sum of the fraction of all pairs of shorter paths that go through v , defined in (1), as follows:

$$cB(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \quad (1)$$

Formula for calculating the Betweenness centrality measure

B. “HITS”

Hits presents two key concepts to perform connection

analysis, these are Hubs and Authorities. "Hubs are vertices that point to many of the other vertices that are considered important. Authorities are those important vertices. From here comes a circular definition: good hubs are those that target many good authorities and good authorities are those targeted by many good hubs" [6]:

1) Hub

A Hub measures the intensity at which the vertices point to a vertex [6]. In other words, they are sources of information which are very likely to be used by other vertices. For that matter, the intensity with which the subjects are a prerequisite for others.

2) Authority

Authority measures the intensity with which one vertex is pointed by many vertices [6]. In this case, the intensity of the prerequisites of a subject. That is, the intensity of how the subjects use the information from other subjects (possibly hubs).

3) Relationship to case study

The HITS algorithm [6] is suitable for the case study for multiple reasons:

- The way the algorithm iterates takes into account not only the immediate adjacent vertices, but their importance in the entire network configuration. This adapts very well because it allows to see beyond the immediate prerequisites of a subject, analyzing the entire flow of the study program in a complete way.
- The final values of hubs are significant to define which subjects (vertices) are most relevant in that they provide the necessary knowledge for the student to take higher subjects.
- If it has a high Authority value, it means that the knowledge of previous subjects is necessary to be able to take that subject. In other words, it is a subject that condenses knowledge of many subjects.

The individual indicators are the results obtained from executing the previous algorithms on the graph, these indicators are normalized and the average of the 3 results obtained represents the final relevance index.

The normalization is carried out by calculating the proportional weight of the indicator value against the sum of the indicators of all the subjects, according to in (2), as follows:

$$betwetness_norm_i = \frac{betwetness_i}{\sum_{i=1}^n betwetness_i} \quad (2)$$

Formula for normalizing betweenness centrality measures.

C. Methodology validation

Results were validated through statistical analysis using historical records of students about approval or failure in subjects and the number of semesters they took to complete their study plan. This allowed to corroborate the validity of the presented methodology and is shown in full detail in Section IV. Result Analysis.

V. RESULTS

According to what is proposed in the methodology, it can be verified that it is possible to model the graph having the subject codes as vertices, and how the relationships between subjects that have other prerequisites are edges. That is, if vertex A corresponds to Mathematics 1, Vertex B to Mathematics 2, and there is also an edge that goes from A to B, it means that Mathematics 2 has Mathematics 1 as a prerequisite. In the following figure the result of modeling the entire Systems and Computing Engineering program is shown, following the methodology outlined before.

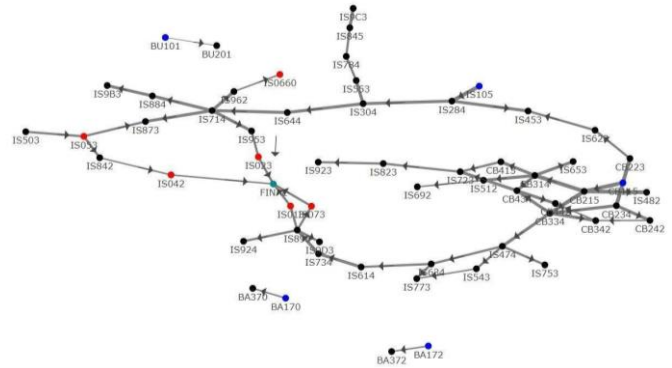


Fig. 1. Graph representation of the curricular grid of the computer and systems engineering program.

In Fig. 1, the representation of the vertices of the graph as the subjects can be seen. The first semester subjects (starting vertices) are represented in blue, the last semester subjects (ending vertices) are represented in red, there is an additional vertex, called End which is used to mark the end of the study plan, the arrows indicate the direction of the study plan, marking the connections between the subjects. Likewise, some lines(edges) are wider than others, representing the number of credits between the subjects.

Tables I, II and III present the results of each of the indicators, showing the subjects with the highest value in intensity measures.

TABLE I
TOP 5 OF SUBJECTS WITH THE HIGHEST BETWEENNESS CENTRALITY INDICATOR.

Code	Name	Betweenness
CB334	Física 2	0,014
IS474	Fundamentos De Electrónica	0013
IS634	Electrónica digital	0,012
IS614	Arquitectura de computadores	0,011
IS734	Sistemas Operativos 1	0,011

TABLE II
TOP 2 OF SUBJECTS WITH THE HIGHEST AUTHORITIES INDICATOR.

Code	Name	Hubs
CB215	Matemáticas 2	0,718
CB234	Física 1	0,282

TABLE III
TOP 4 OF SUBJECTS WITH THE HIGHEST HUB INDICATOR

Code	Name	Authorities
CB334	Física 2	0,36
CB314	Matemáticas 3	0,28
IS482	Teoría General de Sistemas	0,28
CB242	Laboratorio de Física I	0,08

CB334	Física 2	0,096	0	0,362	0,153
CB234	Física 1	0,019	0,28	0	0,1004
CB314	Matemáticas 3	0,024	0	0,275	0,0999
IS482	Teoría General de Sistemas	0	0	0,275	0,092
IS474	Fundamentos De Electrónica	0,090	0	0	0,030
CB242	Laboratorio de física I	0,002	0	0	0,030
IS634	Electrónica digital	0,080	0	0	0,027

The following graph presents the final relevance index created from the weighting of the 3 previous results.

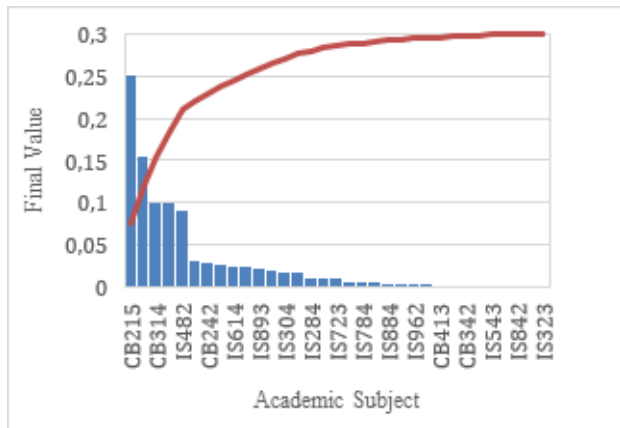


Fig. 2. Pareto chart with the representation of the final index for each subject.

Table IV presents the relevance index, built from the previous indicators, this is made up of the sum of the 3 normalized initial measurements, the top 8 results obtained, corresponding to 80% (approximately) of the relevant subjects.

TABLE IV
SUBJECTS THAT MAKE UP 80% OF THE PARETO MEASURE, WITH THEIR VALUES IN THE INTERMEDIATE INDICATORS AND THE FINAL INDICATOR.

Code	Name	Betweenness	Hubs	Authorities	Relevance index
CB215	Matemáticas 2	0,032	0,718	0	0,247

The process that is described before, was applied to all the engineering programs at the university and because multiple subjects are common to different careers, it is possible to compare the results that were obtained in these across different programs. The obtained results are synthesized Table V.

TABLE V.
RELEVANCE INDEX MEASUREMENT FOR THE ENGINEERING PROGRAMS IN UNIVERSIDAD TECNOLÓGICA DE PEREIRA.

Subject	Academic Program				
	Ingeniería Mecánica	Ingeniería Industrial	Ingeniería de Sistemas y Computación	Ingeniería Eléctrica	Ingeniería Física
CB234	0,46	0,19	0,30	0,38	
CB314			0,30	0,30	
CB334	0,32	0,19	0,46	0,42	0,14
CB215	0,34		0,75	0,66	
IS482			0,28		
IM423				0,38	
CB3A4		0,57			
CB4A3		0,30			

CB4A4	0,22
IM303	0,16
IM533	0,14
CB434	1,12
IF713	0,18
IF513	0,15

The table is composed of rows that contain the most important subjects determined by the study, and in the columns are all the engineering programs. In the intersection between rows and columns, it's the measurement that was obtained for the relevance index in each subject. If the intersection is empty it means that either the subject is not offered by the program, or that the measurement is not significant.

VI. RESULT ANALYSIS

In order to validate the previous results, we proceeded to contrast the subjects that were determined by these measures as relevant with the historical files of graduates and approval of student subjects, in order to demonstrate impacts on academic transit.

The following graphs show the population analysis of the subjects highlighted by the indicators.

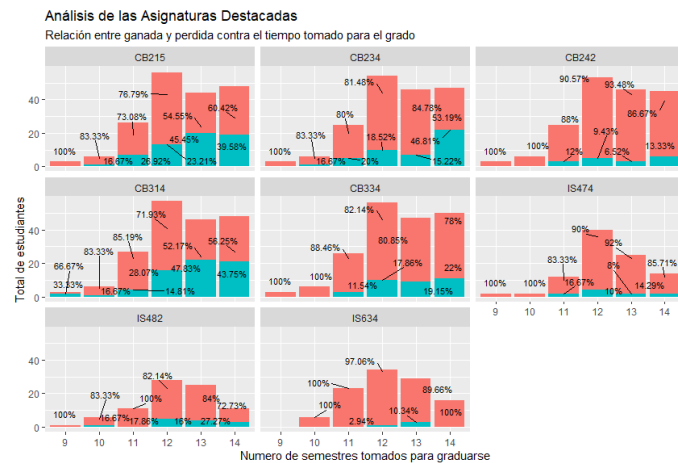


Fig. 3. Loss and gain percentages of the subjects whose relevance index make up 80% of the total.

Fig. 3 shows the relationship between the state of the subject (approved / failed) and the time the student took to graduate. The x-axis shows the semesters used for the completion of the study plan and the y-axis shows the percentage of students who approved or failed it. In general terms, the less these subjects are failed, the less time they take in their study plan. This may mean that by missing this subject, students take longer to complete the curriculum.

In Table V, it can be seen that although the different study plans have common subjects, the relevance index is different

between them, which shows the singularity in the transitions in each of them, that proves the replicability of the methodology.

VII. CONCLUSION

It is possible to represent a study plan of a career by means of a directed graph, and by graph techniques to determine the relevance of their subjects according to their location in the network, their credits and their prerequisites. The proposed index, made up of a weight between scores of Intermediation, Hub, and Authorities, can be very useful when formulating a study plan or analyzing the causes for a student to graduate in a timely manner.

The formulation of this indicator can also be a complement to other analyzes, since it gives a different criterion to what is known of a subject, which currently are only its academic credits and its prerequisites. This not only leads to having a greater knowledge of the content of the curriculum of a program, but also knowing how much relevance each subject has with respect to its contribution in transit through the program. The results allow the proper attention to be paid to the relevant subjects, allowing students to transit through the study plan and its effective completion and also proves the replicability of the methodology as it was used in different engineering program at Universidad Tecnológica de Pereira besides from Systems and Computer Engineering.

Based on the proves and analyzes shown in this paper the proposed hypothesis can be proven, which is that graph techniques can allow determining the relevance of the subjects in a study plan by quantifying the intermediations based on prerequisites and academic credits through the HITS algorithm and Betweenness Centrality measure.

REFERENCES

- [1] S.S. Ray. "Introduction to Graphs", in Graph Theory with Algorithms and its Applications, 2013, Springer, ISBN: 978-81-322-0749-8. Available: <https://link.springer.com/> DOI: 10.1007/978-81-322-0750-4
- [2] S. Guze. "Graph Theory Approach to the Vulnerability of Transportation Networks". Department of Mathematics, Gdynia Maritime University. 12 December 2019. Available: <https://www.mdpi.com/1999-4893/12/12/270/htm>
- [3] C. Ding, X. He, P. Husbands, H. Zha, H. Simon. Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval. Association for Computing Machinery, 2002, New York, USA. Available: <https://dl.acm.org/doi/10.1145/564376.564440>. DOI: 10.1145/564376.564440
- [4] Ulrik Brandes. A faster algorithm for betweenness centrality. Journal of Mathematical Sociology, 25(2):163–177, 2001.
- [5] Neo4j. "The Betweenness Centrality algorithm". Neo4j Labs Graph Algorithms library. Available: <https://neo4j.com/docs/graph-algorithms>
- [6] M. Benzi, E. Estrada, C. Klymkoc. "Ranking hubs and authorities using matrix functions". Linear Algebra and Its Applications, Elsevier, 2012.

Jhoniers Gilberto Guerrero Erazo was born in the municipality of Colon in the Department of Nariño, Sanitary Engineer from “Universidad del Valle”, Master in Sanitary and Environmental Engineering from the same university and Doctor of Engineering Sciences in “Berlin Institute of Technology”.



He worked as a contractor for the Regional Autonomous Corporation of Nariño, consultant of the Basic Sanitation team of “Corponariño” - GTZ-Corponariño, teacher of the Nariño University, at “Universidad Tecnológica de Pereira” has served as head of the Basic Science Department in the Environmental Sciences Faculty, Director of Environmental Administration School, Director of the Environmental Sciences Graduate School, Dean of Environmental Sciences School, Full Professor and since December 2004 he works as Academic Vice-rector of the same College.

ORCID: <https://orcid.org/0000-0002-0243-2320>

Germán Stiven Grandas Aguirre was born in the municipality of Pereira, in the department of Risaralda, Colombia. Computer Science Engineering of “Universidad Tecnológica de Pereira” in 2019. Since 2019, he works as a co-researcher within the same university in areas such as data science, educational data mining and machine learning.



ORCID: <https://orcid.org/0000-0002-1307-2818>

Juan Diego Castaño Gómez was born in the municipality of Pereira, in the department of Risaralda, Colombia. Computer Science Engineering student since 2016 with 2 years of experience as co-investigator in data analytics projects. Dedicated to investigation methodologies, data mining, data analytics and tools for group collaboration and software development.



ORCID: <https://orcid.org/0000-0001-9091-8592>