STUDY OF A BOOK. APPLICATION OF THE GRAPH THEORY

Joseph Maria Franquet Bernis

Agricultural Engineer PhD, EUR-ING., Economic and Business Sciences PhD, Dr. in Architecture (UIC). Diploma in Operational Research. Universidad Nacional de Educación a Distancia (UNED). Northeast Campus. Associated Center of Tortosa (Tarragona, Spain). director@tortosa.uned.es.

SUMMARY / ABSTRACT

Graph theory is, no doubt, an important part of operational research, productive applications in economy and technique, although, as we will see next, you can also create a wide field of utilities in pedagogy and Education Sciences. In this work these concepts will be applied to the efficient study of a book. In the example from the book here we develop, the vertices are the different chapters of the same and refers to bows to denote the time needed to carry out the activities of study and understanding necessary to properly assimilate a particular chapter incident. Finally, it is performed the temporal weighting of the graph that offers, as a result, the obtaining of the paths of minimum and maximum duration in the process of study and assimilation of the text.

Key words: graph, vertex, arc, activity, weighting, minimum and maximum path algorithm.

RESUMEN

La Teoría de Grafos constituye, sin duda, una parte importante de la Investigación Operativa, de fecundas aplicaciones en la Economía y en la Técnica, aunque, como veremos seguidamente, también puede crear un extenso campo de utilidades en la Pedagogía y en las ciencias de la educación. En este trabajo se aplicarán estos conceptos al estudio eficiente de un libro. En el ejemplo del libro que aquí desarrollamos, los vértices son los diferentes capítulos del mismo y se entiende que los arcos denotan el tiempo necesario para llevar a cabo las actividades de estudio y comprensión necesarias para asimilar correctamente un capítulo determinado incidente. Se lleva a cabo, por último, la ponderación temporal del grafo lo que ofrece, como resultado, la obtención de los caminos de duración máxima y mínima en el proceso de estudio y asimilación del texto.

***Palabras clave**: grafo, vértice, arco, actividad, ponderación, camino máximo y mínimo, algoritmo.*

INDEX

Page

* * * * *

I. INTRODUCTION

A "graph" is the representation, by means of sets, of arbitrary relations existing between various objects. His theory constitutes a basic tool of Operations Research[1]. There are two types of graphs depending on whether the relationship between the objects is univocal or biunivocal (bijective). The former form directed graphs or digraphs and the latter non-directed graphs or simply graphs. In most of the algorithms under study, reference is made to the basic terminology proposed below. Such terminology, however, is not standard and may end up varying in the different texts that can be found on the subject.

In mathematics and computer science, graph theory (also called graph theory) studies the properties of graphs (also called graphs). A graph is a non-empty set of objects called vertices (or nodes) and a selection of pairs of vertices, called edges (*arcs* in English) that can be oriented or not oriented. Typically, a graph is represented by a series of points (the vertices) connected by lines (the edges).

[1] Operations research or operational research (also known as decision-making theory or mathematical programming) (OR) is a branch of mathematics that involves the use of mathematical models, statistics, and algorithms to perform a rational decision-making process. Frequently, it deals with the study of complex real systems, in order to improve (or optimize) their operation. Operations research allows analysis of decision-making taking into account the scarcity of resources, to determine how a defined objective can be optimized, such as maximizing benefits or income, or minimizing costs.

II. METHODOLOGY

1. Basic definitions

Graph theory is one of the theoretical parts of mathematics in which the notion of "multivocal correspondence" is very useful, that is, when there is an element of the initial set with more than one image. Well, now consider a finite set $V = \{v_1, v_2, \cdots, v_n\}$ and a multivocal correspondence Γ defined on this set. The pair G = (V, Γ) is said to constitute a graph of order n, which can be represented with the help of a drawing called "sagittal graph representation". Each element v_i is made to correspond to a point on the paper, called the "vertex" of the graph. Two vertices v_i and v_j that are linked by an arrow from v_i to v_j are called adjacent. This arrow, called the graph arc, represents the relationship between the two elements v_i and v_j of set V (Desbazeille, 1969).

A graph is said to have no loops when the main diagonal of the matrix associated with it contains nothing but zeros. When $(v_i, v_i) = 1$, a loop is said to exist at vertex v_i.

Let $a = (v_i, v_j)$ be any arc of the graph G. The vertex v_i is called the *initial extremity* of the arc and the vertex v_j its *terminal extremity*. It is also said that a is an arc incident internally to v_j and incident externally to v_i. The *interior or exterior degree* of a vertex is the number of arcs incident inside or outside this vertex.

An ordered sequence of arcs $(a_1, a_2, ..., a_p)$ is called the *path* such that the terminal extremity of each arc coincides with the initial extremity of the following arc. When the terminal extremity of the last arc is confused with the initial extremity of the first arc, the (finite) path forms a *circuit*. Unless otherwise indicated, the length of a path or circuit is equal to the number of arches that make it up. When these arcs are all different, the path or circuit is said to be *simple*, and when they have vertices (all of which are extremities), they are all *elemental*.

A path or a circuit that passes once and once for each vertex of the graph is called a *Hamiltonian*. Such a path or circuit can be characterized by the following double property: being elementary and of length *n*, in the case of a circuit, or of length *n-1* in the case of a path, where *n* is the order of the graph.

The method of Latin composition presented by A. Kaufmann and Y. Malgrange in the *Journal of the French Society for Operational Research*, VII, number 26, edited by Dunod (cited in the bibliography) is used to advantage in the search for Hamiltonian roads and circuits, which we cannot expose here due to lack of space.

Here are different types of graphs that have particular properties, namely (Franquet, 2008):

- *Symmetric graph*: in which two adjacent vertices are always linked by two arcs (one in each direction):

$$(v_i,v_j) \in A \Rightarrow (v_j,v_i) \in A;$$

- *Antisymmetric graph:* in which two vertices are never linked by two arcs:

$$(v_i,v_j) \in A \Rightarrow (v_j,v_i) \notin A;$$

- *Complete graph*: in which any two vertices are always adjacent; or in other words, that all pairs of vertices are linked in at least one of the two directions;

- *Strongly connected graph*: in which two different vertices are always linked by at least two paths (one in each direction);

- *Transitive graph*: in which there is always an arc that goes from the origin of any path to its extremity; in addition, each vertex has a loop;

- *Graph without circuits*: in which there is no circuit, not even a loop;

- *Simple graph*: in which there is a division of two vertices into two classes in such a way that every arc has its initial extremity in the first and its terminal extremity in the second. A simple graph is often expressed like this: G = (V, W, Γ).

A "directed graph" or "digraph" consists of a set of vertices V and a set of arcs A. The vertices are called *nodes* or *points*; arcs are also known as *edges* or directed *lines* that represent that, between a pair of vertices, there is a univocal relationship aRb but not necessarily bRa (in which case there would be a “circuit” between those nodes). So arcs are commonly represented by ordered pairs (a, b), where a is said to be the head and b is the tail of the arc, and is often also represented by an arrow, as shown in the following figure:

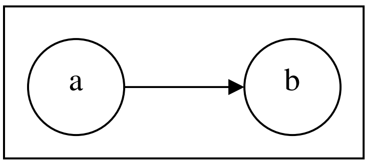

FIG. 1. Directed graph.

The graph can also be defined as: $G = \{V, A\}$ where $V = \{v_1, v_2, \cdots, v_n\}$, $A = \{a_1, a_2, \cdots, a_n\}$ and $a_i = (v_j, v_k)$ such that $v_j, v_k \in V$. In this graph it is understood that $(v_i, v_j) \neq (v_j, v_i)$ and in many cases there is only one of the pairs of vertices.

A vertex that only has arcs coming out of it is called a *source*, and a vertex that only has arcs directed toward it is called a *sink*. Such nomenclature is important when digraphs are used to solve flow problems.

An undirected graph, or graph, like a digraph, consists of a set of vertices V and a set of arcs A. The difference between the two is that the existence of aRb presupposes that bRa also exists and that they are the same. Thus, it is indistinct to speak of the arch (a, b) or (b, a), just as it does not make sense to speak of the "head" or "tail" of the arch. These graphs are represented schematically as indicated in figure 2, where the circles represent the vertices and the lines represent the arcs. So:

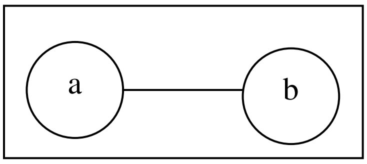

FIG. 2. Undirected graph.

In the latter case, $G = \{V, A\}$ where $V = \{v_1, v_2, \cdots, v_n\}$, $A = \{a_1, a_2, \cdots, a_n\}$ and $a_i = (v_j, v_k)$ such that $v_j, v_k \in V$. In this graph it is understood that $(v_i, v_j) \Leftrightarrow (v_j, v_i)$ and it is also true that: $(v_i, v_j) = (v_j, v_i)$, where both pairs of vertices represent precisely the same arc.

There are also graphs where the arcs have some value associated (in our case it could be the foreseeable time of assimilation by the reader of a certain chapter of the book), in which case we speak of "weighted graphs" and now they are represented arches like triplets. Therefore, there is still information on the vertices joined by said arc, in addition to information on the weight or weight of said arc or activity. Thus, the arc is represented as: $a = (v_i, v_j, w)$ where v_i, v_j are the origin and destination and w is the weight (the time expressed in minutes, in our case), respectively.

Using the new terminology, then, let's see that a node **b** is said to be **adjacent** to node **a** if the arc (a, b) exists. Note that, for an undirected graph, **a** is necessarily also adjacent to **b**. This does not occur in directed graphs where the existence of (a, b) does not imply that (b, a) also exists. This concept is of particular importance, since graphs are usually represented on the computer by means of lists or matrices of adjacencies.

An arc (a, b) **affects** node b, just as in an undirected graph said arc also affects node a because arc (b, a) also exists. The number of arcs that impact a node gives the **degree** to that node. The node with the highest degree in the graph indicates the degree of said graph. It is also customary to represent a graph through incident lists or matrices (Franquet, 2013).

2. Leveling of the graph

2.1. Conceptualization

When dealing with the manual construction of the graph of a book or study, it is very useful to order the activities by levels. The ordering by levels allows the graph in question to be constructed by arranging the events in such a way that when plotting the activities or priorities, an excessive number of crosses does not appear, which would make it difficult to interpret the graph in the book. In the example of the book that we develop here, the vertices will be the different chapters of the same in number of nine (or ten, considering a hypothetical introductory chapter), and it is understood that the arches denote the time necessary to carry out the study activities and understanding necessary to correctly assimilate a certain incident chapter, a question that we will see later, which will ultimately result in the following graph:

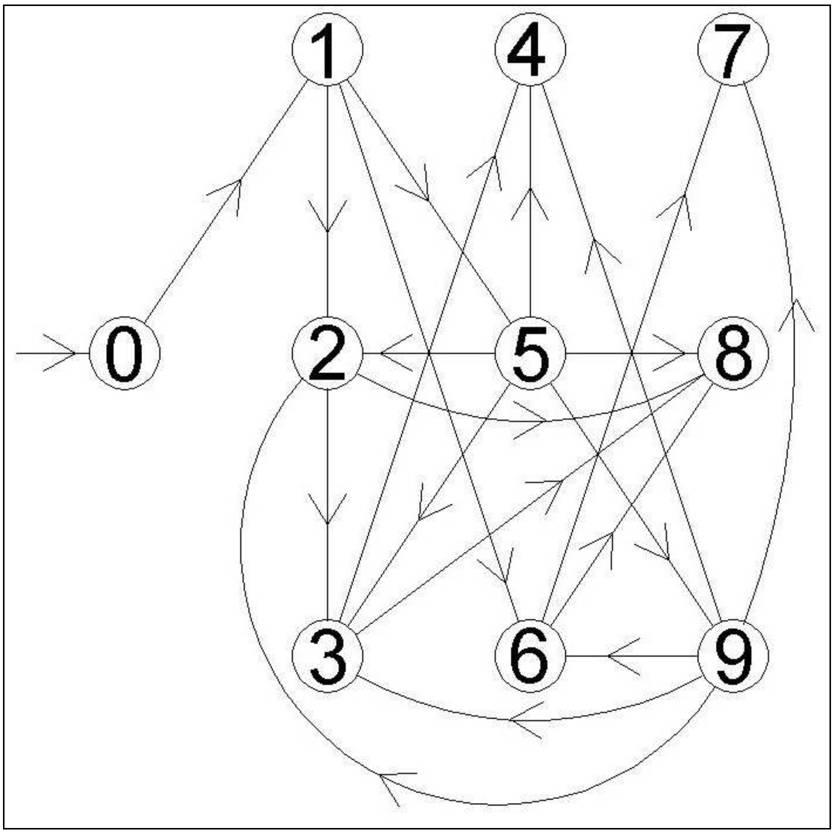

FIG. 3. Graph of the book.

Here the decisive intervention of the reader of the book must take place, or better still of the author himself, who must establish the priorities existing between the different chapters or parts of the book in order to be able to construct the graph in question.

2.2. Graphic method

To do this, the following steps must be followed:

1.- The graph is searched for the subset of vertices from which no arc arises. This subset constitutes the last level of the graph.

2.- Next, we delete these vertices and the arcs related to them.

3.- In the subgraph obtained, the subset of vertices from which no arc is born is searched again. This subset constitutes the penultimate level of the graph.

4.- Next, we eliminate these vertices and the arcs related to them.

5.- Repeating this process iteratively, we obtain the graph ordered in levels.

6.- Note, finally, that in the numbering of the vertices of an activity, the number of the origin event is always less than the number of the final event.

2.3. Matrix method

It is the one that we will use in our case. To do this, the following steps must be followed (Demoucron-Malgrange-Pertuiset algorithm):

1.- Concept of matrix associated with a graph: It is a square matrix of dimension n, equal to the number of vertices, in which its elements a_{ij} are 1 or 0 depending on whether or not there is an arc between vertex i and vertex j.

2.- We enlarge the matrix associated with the graph by means of a certain column vector V_1. The elements of this vector are equal to the sum of the elements of each row of the associated matrix.

3.- The elements of the column that are zeros, indicate the vertices that constitute the last level of the graph.

4.- We expand the associated matrix by a new vector column V_2. The elements of this new vector are obtained by subtracting, from the elements of V_1, the homologous elements of the column (s) that correspond to the vertices that in said vector V_1 take the value zero. When the minuend and subtrand are zero, a cross is placed in the vector instead of zero.

5.- Under the column corresponding to each vector, the numbers of the vertices with which zero value elements are obtained in the vector are placed. The elements of V_2 that are zero will be the vertices of the penultimate level.

6.- Repeating this process iteratively, we obtain the vertices of the other levels, that is, the other column vectors that represent the ordering in levels of the graph, until the last vector appears in which all its components are blades.

As can be seen, this is a connected graph with no circuits. In this way, following the matrix method previously exposed, which leads to the ordering of the vertices in levels towards the antibasis by the method also known as "elimination of descendants", we can form the corresponding Democron-Malgrange-Pertuiset algorithm (Chartrand and Oellermann, 1993), namely:

ALGORITMO DE DEMOUCRON

	0	1	2	3	4	5	6	7	8	9		$\vec{v}_0$	$\vec{v}_1$	$\vec{v}_2$	$\vec{v}_3$	$\vec{v}_4$	$\vec{v}_5$	$\vec{v}_6$	$\vec{v}_7$
0		I										1	1	1	1	1	1	0	X
1			I			I	I					3	3	2	1	1	0	X	X
2				I					I			2	1	0	X	X	X	X	X
3					I				I			2	0	X	X	X	X	X	X
4												0	X	X	X	X	X	X	X
5			I	I	I				I	I		5	3	2	1	0	X	X	X
6								I	I			2	0	X	X	X	X	X	X
7												0	X	X	X	X	X	X	X
8												0	X	X	X	X	X	X	X
9			I	I	I		I	I				5	3	1	0	X	X	X	X
Σ	0	1	3	3	3	1	2	2	4	1		④ ⑦ ⑧ nivel sup.	③ ⑥	②	⑨	⑤	①	⓪ nivel inf.	-

FIG. 4. Demoucron-Malgrange-Pertuiset algorithm.

METHOD:

$$\vec{v}_1 = \vec{v}_0 - ④ - ⑦ - ⑧$$
$$\vec{v}_2 = \vec{v}_1 - ③ - ⑥$$
$$\vec{v}_3 = \vec{v}_2 - ②$$
$$\vec{v}_4 = \vec{v}_3 - ⑨$$
$$\vec{v}_5 = \vec{v}_4 - ⑤$$
$$\vec{v}_6 = \vec{v}_5 - ①$$
$$\vec{v}_7 = \vec{v}_6 - ⓪$$

Now, the ordered graph of the book turns out to be the following:

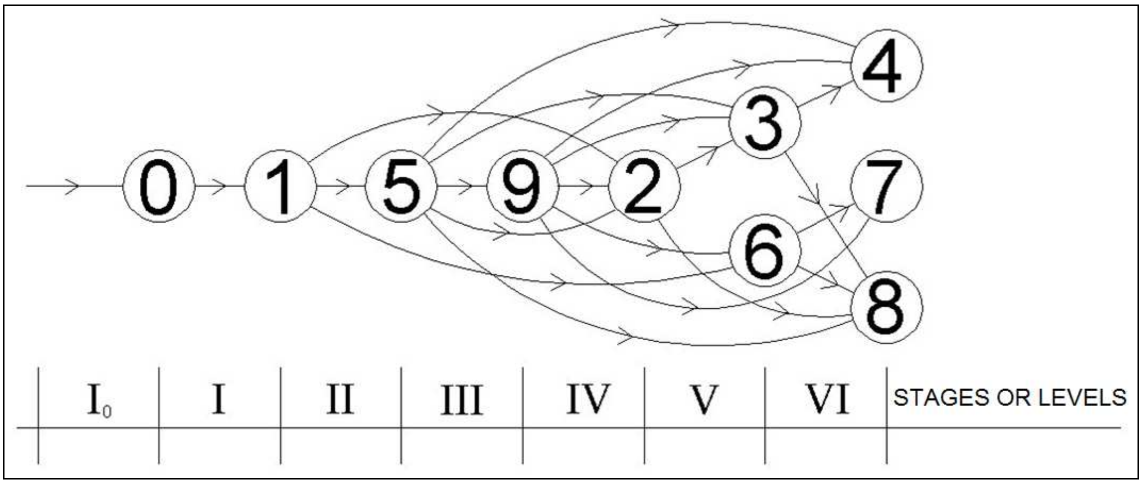

FIG. 5. Graph arranged in levels of the book.

Through the previous arrangement, a clear preference has been revealed between the various stages of the advisable scheme of study and assimilation of the contents of the book. In any case, it must be fulfilled that:

1) All the chapters of the book of the same level must not have "ascendants" in the next level.

2) The order of study of the vertices or chapters of the same level is independent.

3. Time weighting of the graph

Finally, we will give the arcs of the graph their corresponding value expressed, for example, in minutes. The time it takes to carry out an activity is not exactly known, so time estimates must be made. The PERT[2] method considers three different time estimates, namely:

- Optimistic estimate (E_o): it is the minimum time in which the activity could be executed if no undesirable setback arose. In the absence of other determinations, we will set it approximately here at 10 'per printed page, regardless of whether it is complete or not, so it will be given for (10·n) minutes, with *n* being the number of pages of the chapter in question. book.

- Most probable estimate or modal estimate (E_m): it is the time that will be used to execute the activity under normal circumstances; In this case, a value of (25·n) minutes will be assumed.

- Pessimistic estimate (E_p): it is the maximum execution time of the activity if the study circumstances are very unfavourable; in this case, a value of (40·n) minutes will be assumed.

The PERT (D) time will be the weighted arithmetic mean or mathematical expectation of the previous estimates, that is:

$$D = \frac{E_o + 4E_m + E_p}{6} = \frac{10 \cdot n + 100 \cdot n + 40 \cdot n}{6} = (25 \cdot n) \text{ minutes}$$

On the other hand, the variance and/or the standard or "standard" deviation of any activity could also be taken into account, defined as follows:

$$V^2 = \frac{(E_o - E_p)^2}{36} = \frac{(10 \cdot n - 40 \cdot n)^2}{36} = 25 \cdot n^2,$$

being the standard or "standard" deviation of value: V = 5·n. Activities with greater variance obviously have a higher risk of error in estimating their duration.

Here we have assumed a certain number of pages for each chapter of the book (taken, by the way, from a real example), with a total of 714.

[2] The **Project Evaluation and Review Techniques**, commonly abbreviated as **PERT**, is a project management and management model invented in 1958 by the Department of the Navy Department of Special Projects Office of Defense of the USA as part of the Polaris mobile ballistic missile project launched from a submarine. PERT is basically a method of analyzing the tasks involved in completing a given project, especially the time to complete each task, and identifying the minimum time required to complete the entire project. This project model was the first of its kind, a revival for the scientific administration, founded by Fordism and Taylorism. The project model is not very common, since they are all based on PERT in some way. Only the DuPont Corporation's *Critical Path Method* (CPM) was invented at almost the same time as PERT.

Chapter	N	D	V	t(h)	%	% acc.
0	12	300	60	5.00	1.68	1.68
1	34	850	170	14.17	4.76	6.44
2	144	3600	720	60.00	20.17	26.61
3	152	3800	760	63.33	21.29	47.90
4	42	1050	210	17.50	5.88	53.78
5	92	2300	460	38.33	12.89	66.67
6	72	1800	360	30.00	10.08	76.75
7	16	400	80	6.67	2.24	78.99
8	72	1800	360	3.00	10.08	89.07
9	78	1950	390	32.50	10.93	100
TOTAL	714	17850	3570	297.50	100	---

FIG. 6. Activity table.

Let's see the previous table with the number of pages and the duration of each activity D according to the different chapters of the book, as well as their corresponding standard deviation V and the time spent in the assimilation of each one of them and the accumulated since the beginning of the study of the book. Obviously, the Pearson coefficient of variation, which is a relative measure of dispersion of the values of the statistical random variable "activity duration", will be: $\frac{V}{D} \times 100 = 20\%$ in all cases. Note also that the assimilation of all the chapters of the book would involve, according to the assumptions already expressed, a duration of 17850 minutes (exactly a time of 297 hours and a half). It would be possible, however, to achieve the assimilation of the chapters of the last level (if this is of interest) without necessarily going through the study of some others, either by covering routes of maximum or minimum duration, as will be seen below. On the other hand, the arithmetic mean of the duration of each activity (study and assimilation of each chapter) would be 17850/10 = 1785 minutes = 29.75 hours/chapter, and 2/3 of the complete study of the text would be reached at the end of the Chapter 5.

In short, under these temporal conditions, the graph of the book in which the paths of maximum value have been searched and the fictional chapter O' has been added, will be configured as follows:

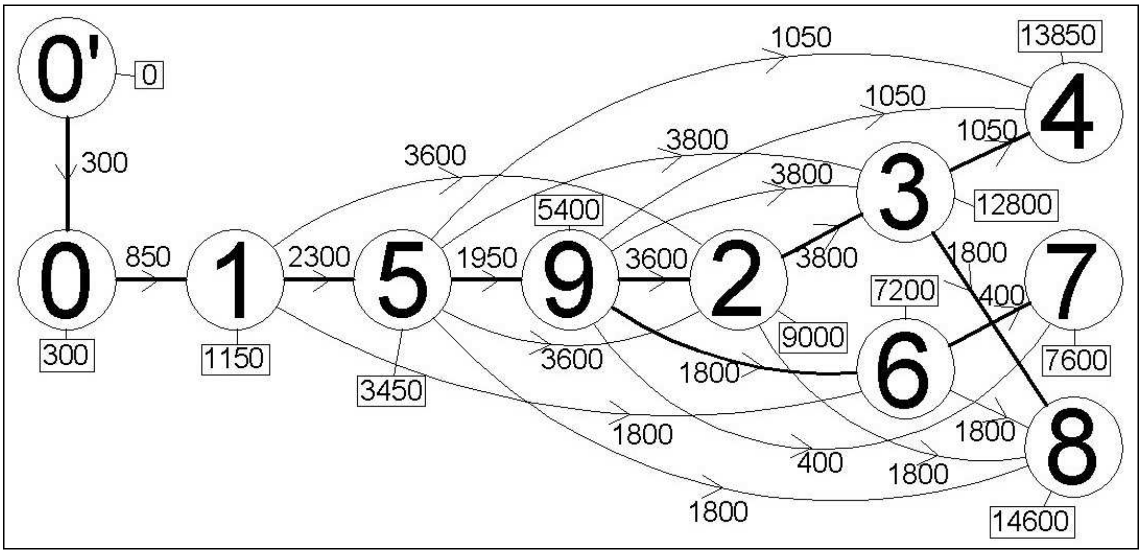

FIG. 7. Time-weighted graph of the activities. Maximum path.

Thus, three paths of maximum length or duration have been obtained, to respectively reach the chapters of the last level or stage, the trace of which has been conveniently highlighted in the previous figure, namely:

[0’, 0, 1, 5, 9, 2, 3, 4] = 13850 minutes (Chap. 4)
[0’, 0, 1, 5, 9, 6, 7] = 7600 minutes (Chap. 7)
[0’, 0, 1, 5, 9, 2, 3, 8] = 14600 minutes (Chap. 8)

In the same way, the graph of the book in which the minimum value paths have been searched, will be configured as follows:

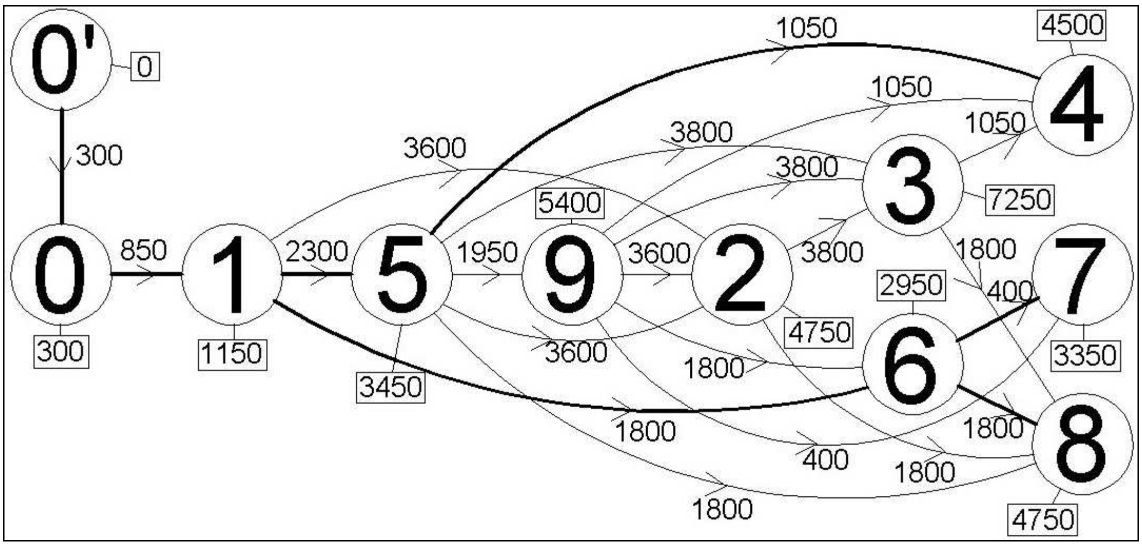

FIG. 8. Time-weighted graph of the activities. Minimum path.

In our case, a simplified version of the *Dijkstra algorithm*, also called the *minimum path algorithm*, could be used. It is an algorithm for determining the shortest path given a vertex origin to the rest of the vertices in a graph with weights on each edge. Its name refers to Edsger Dijkstra, who first described it in 1959[3].

Three paths of minimum length or duration have been obtained, respectively to reach the chapters of the last level or stage, the outline of which has been conveniently highlighted in the previous figure, namely:

[0', 0, 1, 5, 4] = 4500 minutes (Chap. 4)
[0', 0, 1, 6, 7] = 3350 minutes (Chap. 7)
[0', 0, 1, 6, 8] = 4750 minutes (Chap. 8)

What has been stated up to here, for example, would also be applicable to the study of the penultimate level chapters (Chapters 3 and 6) or of any other taking into account, in each case, the path of duration most convenient to the interests of the reader.

III. ELEMENTARY TIPS FOR THE BOOK'S STUDY

At this point, and once the different chapters of the book have been arranged in levels or stages, as well as the different alternatives or itineraries of their assimilation have been presented, I would like to suggest to our readers some ideas about how to focus more efficiently on the study and understanding of a monograph, report, etc. And so we will try to:

- Increase the speed and efficiency of reading:

1. It is about learning to read quickly using the right techniques that allow you to read more and memorize more content in less time, and get more out of what has been read. Some of the skills necessary for a good reading are the following:

- Ability to read and understand at high speeds,
- Ability to use a variable rhythm depending on the purpose and difficulty of the topic,
- Ability to understand the main ideas or central thoughts of the reading material,
- Ability to understand and retain details, good overall retention,
- Ability to appreciate the organization of the material,
- Ability to read critically and evaluative.

2. Ineffective readers read everything at the same speed, while effective readers read three to five times faster and understand the main ideas much better.

[3] The idea behind this algorithm is to explore all the shortest paths that start from the origin vertex and that lead to all the other vertices; when the shortest path from the origin vertex to the rest of the vertices that make up the graph is obtained, the algorithm stops. The algorithm is a specialization of the uniform cost search, and as such, it does not work in graphs with negative cost edges (since always choosing the node with the smallest distance, nodes that in future iterations would lower the general cost of the road when passing an edge with a negative cost).

- Improve concentration:

1. Avoid external and internal distractions.
2. Find a suitable place of study.
3. Eliminate the interruptions raised.
4. Eliminate sound distractions like noise or music with songs.
5. Find the most favourable time for study.
6. Set goals for when to start, stop, and finish.
7. Control mental concerns.
8. Rest periodically 10 minutes every 50 of reading or study.

- Establish the right environment, dedicate the stipulated time, take care of your sight, etc.

IV. CONCLUSIONS

Graph Theory can also be useful in creating a wide field of utility in Pedagogy and in the educational sciences. In this work its methodology is applied for the efficient study of a book. According to the objective of assimilation of knowledge that is pursued in each case, their techniques will allow us to choose the most appropriate and least expensive itinerary in the study and assimilation of the book in question, being able to save the student a lot of time and effort.

BIBLIOGRAPHICAL REFERENCES

1. CHARTRAND, G. & OELLERMANN, O. R. "Applied and Algorithmic Graph Theory". Ed. McGraw-Hill, 1993.

2. DESBAZEILLE, G. *Ejercicios y problemas de Investigación Operativa*. Ed. ICE. Selecciones de Economía de la Empresa. Madrid, 1969. 358 p.

3. DIJKSTRA, E. W. *Go To Statement Considered Harmful*, Communications of the ACM, Vol. 11 (1968) 147-148.

4. FRANQUET BERNIS, J.M. *El estudio operativo de la Psicología: una aproximación matemática.* Ed. Centro Asociado de la UNED. Cadup-Estudios. Tortosa, 2008. 370 p.

5. FRANQUET BERNIS, J.M. *Ecuaciones diferenciales ordinarias y en diferencias finitas. Curso práctico.* Ed. Centro Asociado de la UNED. Cadup-Estudios. Tortosa, 2013. 750 p.

6. KAUFMANN A. & MALGRANGE Y. Revista de la Sociedad Francesa de Investigación Operativa, VII, no. 26. Ed. DUNOD.

<u>LIST OF FIGURES:</u>