ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20240927002023
Original scientific paper

Network and Radial Layouts for Visual Representation of Knowledge Graphs for Cultural Heritage Conservation-Restoration Activities

Álvaro RODRÍGUEZ MIRANDA*, Jaione KORRO

Abstract: This study analyses the spatial arrangement of a knowledge graph documenting conservation-restoration activities of cultural heritage, presented in the case of the monastery of San Millán de la Cogolla (Spain). The graph consists of three groups of nodes, concerning: (1) the physical reality, (2) conducted actions of conservation-restoration and (3) documentary archive. The goal is to explore how tailored spatial layouts can reveal beneficial links for decision-making, and how the analysis of the established paths helps to verify which relevant questions can be managed by the graph. The methodology involves comparing an initial network representation with a radial layout, assessing their complementarity and respective advantages. A novel contribution is that the visual inspection is extended by means of the adjacency matrix, which allows counting all possible paths of any length between nodes, and thus permitting more in-depth analyses about the number of indirect relationships and their meanings. The study highlights the benefits of combining different spatial layouts and computational analyses, integrating the multifaceted nature of graphs (communicative, semantical, and mathematical).

Keywords: conservation-restoration; cultural heritage; knowledge graph; visualization

1 INTRODUCTION

Knowledge graphs are widely used to represent systems in which relationships play a central role. Particularly, when dealing with directed graphs, reference is made to *multivariate networks* [1, 2] wherein, alongside vertex/nodes and links/edges, both attributes exist for the vertices and for the connections.

The conceptual model is the underlying structure of a graph database, i.e., a database populated with items corresponding to each type of entity of the model. For instance, the next conceptual model (figure 1) illustrates the management of the conservation-restoration activities carried out by an organization in charge of a cultural heritage element (e.g., a monastery) [3].

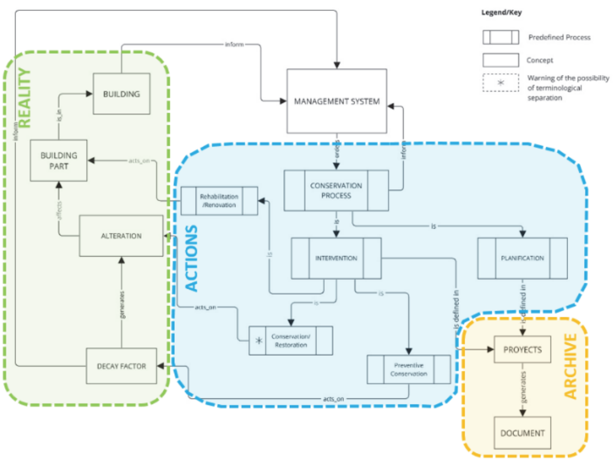

Figure 1 Conceptual model of the activities of conservation-restoration of a cultural heritage element. The different entities of the model fall into one of the following groups: (1) the physical reality of the element, (2) the activities of conservation-restoration carried out by the organization in charge of the element, and (3) the documentary archive generated

The model is based on different proposals in the field of cultural heritage conservation-restoration [4-8], the revision of standards, in particular the CIDOC-CRM [9] i.e., the Conceptual Reference Model of the International Council of Museums (ICOM) International Committee for Documentation and its extensions, notably: CRMba (for archaeological buildings), CRMpe (for research infrastructures), CRMsci (for scientific observations), CRMcr (for conservation-restoration), and Linked Conservation Data (LCD) [10], which facilitates the interconnection with these and other models [11]. Here, the entities are divided into three groups:

1. The physical reality includes the nodes that represent the parts of the building (e.g., church, cloister, refectory, etc.), the alterations (e.g., water stains, cracks, biological colonization, etc.) and the decay factors (e.g., pollution, lighting, water, human action, etc.); as well as the relationships between all these entities.
2. The organization in charge of the upkeep of the monument carries out different types of conservation-restoration actions: (a) planning, (b) preventive conservation works - actions aimed at monitoring and controlling the decay factors that cause the alterations, (c) curative conservation and restoration works - actions to mitigate or correct existing alterations, and (d) renovation/rehabilitation - which acts over the parts of the building to recover a former function, to accommodate new types of use or to comply with new regulations about comfort, security, access, etc.
3. The documentary archive generated by the actions of conservation-restoration.

In particular, a graph database was obtained from the application of this conceptual scheme to the monastery of San Millán de la Cogolla, de Yuso, in La Rioja, Spain. The original reason of the database was to put in order a document collection scattered in four archives and provide a means to contextualize, interlink and, ultimately, recover the information contained therein.

Graphs rely on a mathematical basis (connectivity, size, distances, etc.) [12], but they also have a graphical counterpart. The use of visual variables helps increase the amount of information conveyed in the images; however, with the increasing in size of the graph, the graphical representations become jumbled (hair-balls) [13] and the elements crowd together in such a way that they are impossible to distinguish visually. This fact implies that graphs may be nice but, in many real cases, their practical utility is limited. In response to this situation, some lines of research are dealing with transformations to show large networks effectively and the development of tools for the interactive management of the datasets (through filters, zooms, etc.) [14, 15]. Moreover, customized options are also considered for different user profiles (expert, administrative, casual, etc.) [16]. Finally, the own concept of "complexity" deserves attention, i.e., the mental effort that is necessary to understand and use the information regardless of the number of elements that are presented [17].

There are many options for displaying the elements of a knowledge graph, following different spatial arrangements [18, 19]. For instance, networks use algorithms that combine aesthetic and proximity criteria, considering characteristics such as: (1) giving similar length to all arcs, (2) making the vertices cover all the display area, or (3) reducing the number of crosses between arcs [20]; a prominent example is the graph that shows the links between jazz musicians (https://linkedjazz.org/) developed by the Pratt Institute School of Library Information Science of New York with information extracted from interviews with musicians and other actors related with the world of jazz, which also provides access to samples of the performances. In other cases, however, the homogeneous distribution of the nodes all over the two-dimensional area is replaced by radial charts (on a circle) or by series of parallel rows or columns (these types of representations are called alluvial diagrams, parallel coordinates plots, etc.). In any case, beyond the aesthetic quality, which is always important, some configurations are more suited for the accomplishment of certain tasks than others. An excellent example of application to cultural heritage is the UNESCO's graph for intangible heritage (https://ich.unesco.org/dive/), which shows different ways of visualizing around 500 listed elements, with the possibility of exploring the dataset through domains, themes, geographic location, ecosystems, risks and relationships with the United Nation's sustainable development goals (SDG) for 2030.

After the spatial layout of the nodes and links has been selected, the transmission of the message is completed with the visual variables that, in its original conception [21], are the changes in size, value, colour, texture, shape and orientation. From that initial proposal, many extensions and modifications have been suggested and applied to different fields such as cartography [22] or data visualization [23]. Indeed, the selection of a suitable representation is driven by the "tasks" identified, i.e., by the questions that the data analysis must answer or the activities that the users want to do through the interaction with the visual representation [24]. Questions are divided in "elementary" - referring to information on individual elements of the graph (such as the value of an attribute or connection) and "synoptic" - which are related with the complete set of elements and their behaviour [23].

In essence, the exploration of the knowledge graph matches different perspectives: firstly, semantics permit the exploration using computer programs that move through connections; secondly, graphs can also be seen as communicative elements based on multiple display options and, thirdly, they are mathematical structures allowing the use of metrics and computations.

Turning to the case study, the graph is initially presented as a network; then an alternative radial visualization will be proposed, and the set of connections of the conceptual model will be further developed, thanks to the mathematical formulation provided by the adjacency matrix, with the aim of exploring to which extent searches of variable depth (understanding this concept as the necessary number of steps between entities to answer a query) can be defined. The discussion that follows deals with the complementarity of the visualizations and the central role that the structure of the graph plays concerning its potential to perform tasks. Finally, the conclusions advocate for an integrated consideration of the different aspects of graphs (semantical, communicative, and mathematical).

The novelty of this research is twofold: on the one hand, it stresses the importance of adapted displays for human users, even if the graph database is originally aimed at machine processing, in order to complement the analysis with tasks that may be more effectively completed by users, such as visual inspection and discovery. On the other hand, it is argued that the usefulness of path analysis depends not so much on the considered length, but on a smart design of the conceptual model behind the graph, which takes into account the meaning of the connections.

2 OBJECTIVES

The objective of this paper is to show how different layouts of a graph database highlight specific connections and groups of nodes. Therefore, the design and selection of layouts can assist to better transmit the information contained in the graph, facilitating the navigation, searches, and the rest of tasks to be carried out.

Moreover, to understand user's visual tour on the graph during the accomplishment of tasks, it is important to analyse the number and corresponding meaning of all direct and indirect links. To this end, the adjacency matrix, a tool that quantifies the relationships between elements within a graph, is examined, demonstrating its applicability while also addressing its limitations.

3 GRAPHICAL REPRESENTATION OF A GRAPH

The conceptual model presented in Fig. 1 was filled in with the information concerning the conservation-restoration works carried out in the monastery of San Millán de la Cogolla since its declaration as World Heritage site by the UNESCO in 1997. The goal was to provide an efficient tool (the graph database) to store and access the information about these activities. Altogether, during the considered period, there have been 40 projects that contain 112 documents of different typologies (reports, studies, certifications, notices, etc.).

Regarding the physical reality of the monastery (i.e., parts of the building, deteriorations and decay factors), on the one hand, the monastery was divided in 14 areas according to historical-functional criteria, which have proven useful since they give a clear overview of the spaces and, besides, because they coincide with the separation used in the conservation-restoration actions. Moreover, widely recognized references were selected for the identification of the set of deterioration patterns and the decay factors that can trigger them, in particular, the list of 33 types of alterations listed by ICOMOS [25] and the 10 decay factors considered by the Canadian Conservation Institute [26].

In the following visualization of the resulting graph (Fig. 2), the largest nodes represent the conceptual entities, while the smallest ones are the specific elements of the case study.

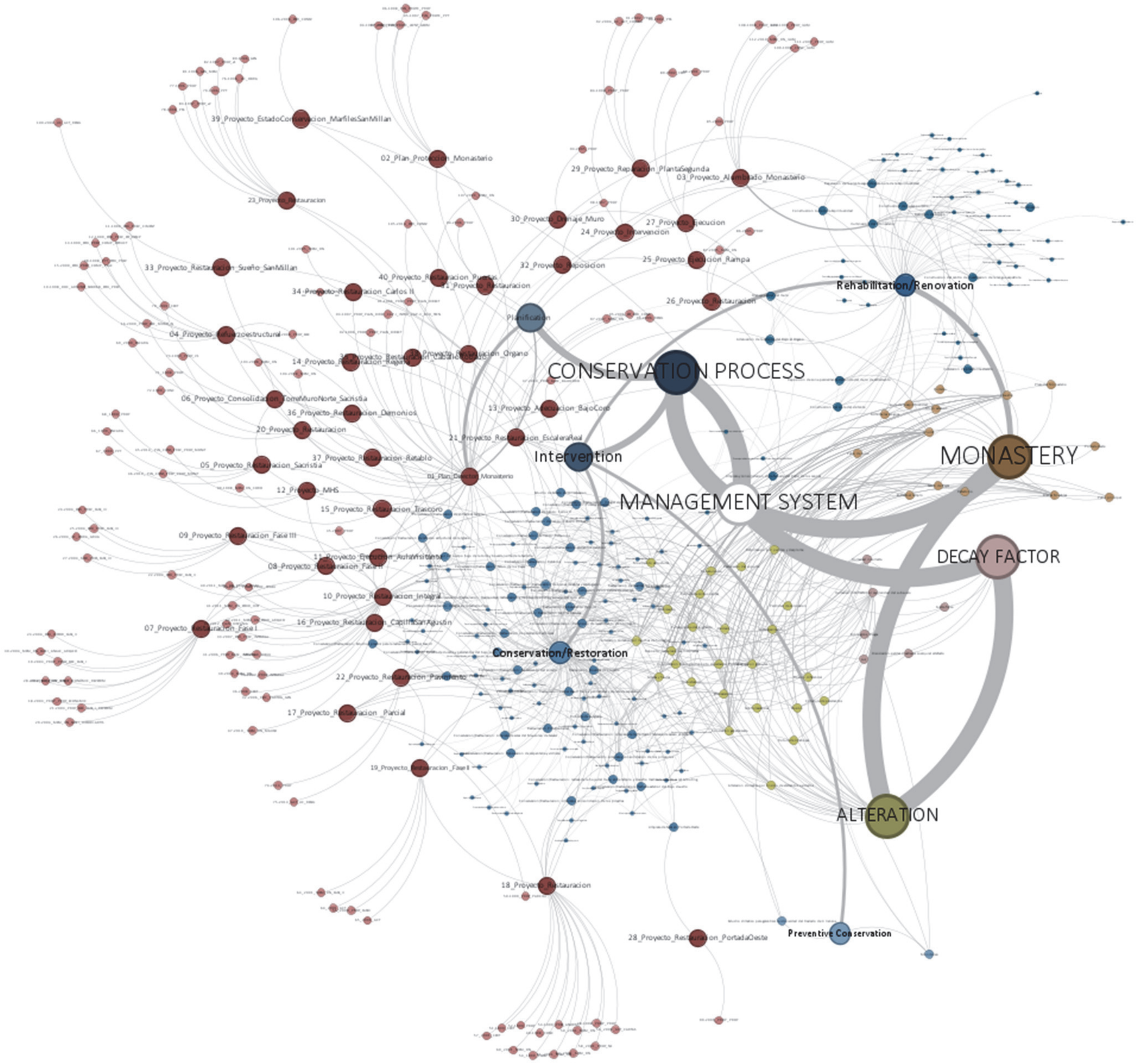

Figure 2 Resulting knowledge graph after applying the conceptual model to the case study of the monastery of San Millán de la Cogolla, de Yuso

The use of visual variables allows a better identification of the type of entities shown. This graph is prepared for static viewing; however, it can be also explored interactively, by selecting and filtering specific nodes with the software employed for its development - Gephy. Nowadays, an interactive web version which includes the possibility of updating the database, interactive navigation and filtering is in the pipeline.

In this layout, clusters of interlinked entities emerge, following the rule that the more interconnected the nodes are, the nearer they appear in the graphical display. For example, it cannot be understood that nodes for two different parts of the building that appear close together in the display are related to spaces that are physically adjacent; by contrast, if they are neighbours it is because they have many relationships in common, such as similar deterioration patterns or the fact of having been considered by the same conservation-restoration works. Likewise, what can be found within the vicinity of a node representing a type of alteration are the nodes concerning factors that can trigger it, the parts of the building where the alteration is present or the conservation actions that have been carried out to mitigate this decay.

Among the contributions of this graph to the management of the monument, it is worth highlighting that it brought to light some documentary gaps (since it clearly shows when the links between "projects" and "documents" are scarce), examples of good practices (where the generated clusters have a dense set of connections) and some others that have not been so good (the ones that generated loosely connected groups). Moreover, the graph revealed the trend towards corrective actions (i.e., the ones focused on the mitigation of the alterations), consistently favoured over the preventive measures, which is the opposite tendency to current criteria for the management of the cultural sites. Likewise, it helped to identify the parts of the building with the largest number of works, the most predominant decay factor at the origin of the main conservation issues of the monastery (the humidity), as well as the deteriorations that are active in the different parts of the building.

Despite those benefits, the network display is not adequate to visually perceive the answers to all possible queries to the database, neither is it intuitive enough for the interactive interpretation by the personnel in charge of the management of the monument. Moreover, there are meaningful relative orders for some of the entities (for instance, the time sequence of the projects carried out on the monument) that are not considered in the network, but which may be worth showing. Hence, complementary visualizations are under analysis.

Consider the following radial layout (Fig. 3). The "physical reality" (i.e., nodes regarding the "parts of the building", "deterioration patterns" and "decay factors") is located on the semi circumference on the left. As can be seen, the spatial arrangement accepts the use of concentric rings to add a hierarchy, which can be useful to include different degrees of granularity. For example, the ICOMOS glossary describes 33 types of "deteriorations" classified in 5 categories (crack & deformation, detachment, features induced by material loss, discoloration & deposit, and biological colonization), therefore, depending on the level of detail of the visualization it is possible to represent this sector only by the top categories or by the complete list of deteriorations. If necessary, the other two sectors may also be considered in a similar way, i.e., by defining hierarchies for the "parts of the building" and for the "decay factors".

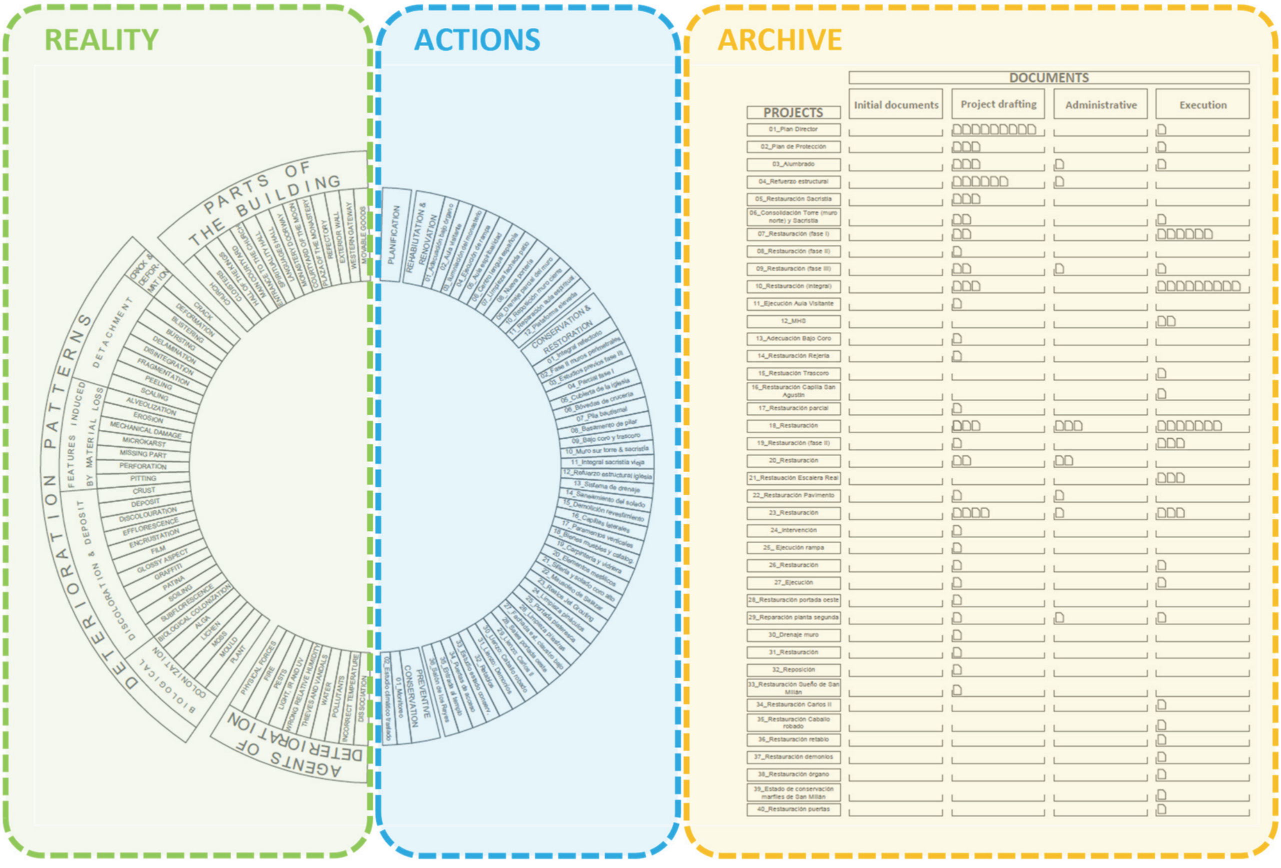

Figure 3 Radial layout of the nodes (entities) of the monastery of San Millán de la Cogolla, de Yuso; with the indication of the three groups of entities

On the right side of Fig. 3, the "documentary archive" provides significant information about the number and distribution of the available documents, giving an idea of the completeness of the documentation for each project. The chart does not only show the number of documents per project, but the documents are also classified according to the phase in which they were created: initial documents - those that justify the commencement of an intervention, such as a notice about the detection of a new deterioration pattern which may endanger the stability of a part of the building, documents related to the writing of the project, definition of the intervention to be carried out, administrative documents - concerning the processes and communication between the agents during the commission of the works, and documents describing the conservation-restoration works. Although, a good documentary register of a project should contain documents in each of these phases, Fig. 3 shows that this is not the common situation, being particularly striking the absolute absence of initial documents in the archive. This deficiency is explained because the need of intervention may seem evident for the organization when an action is commissioned, therefore no document to record it is deemed necessary at that moment; however, this is a mistake, since future users who will be interested in the analysis of past interventions will not have the necessary context to understand them properly.

The link between the physical reality and the documentary archive is established through the "actions of conservation-restoration" (semi circumference on the right, in the middle of the Fig. 3) carried out by the organization.

The set of relationships is represented in Fig. 4. It is important to note that the connections between elements inform about the decay processes that have been detected and are under treatment; nevertheless, the absence of connections does not necessarily imply that some parts of the building are free from problems or that some types of deteriorations are not present. In line with the semantics of the database that is represented, the objective is to reflect the activities carried out by the organization that manages the monument, as well as the level of knowledge about the state of the building. The approach is dynamic, because the deteriorations evolve over time and new actions are carried out to deal with them.

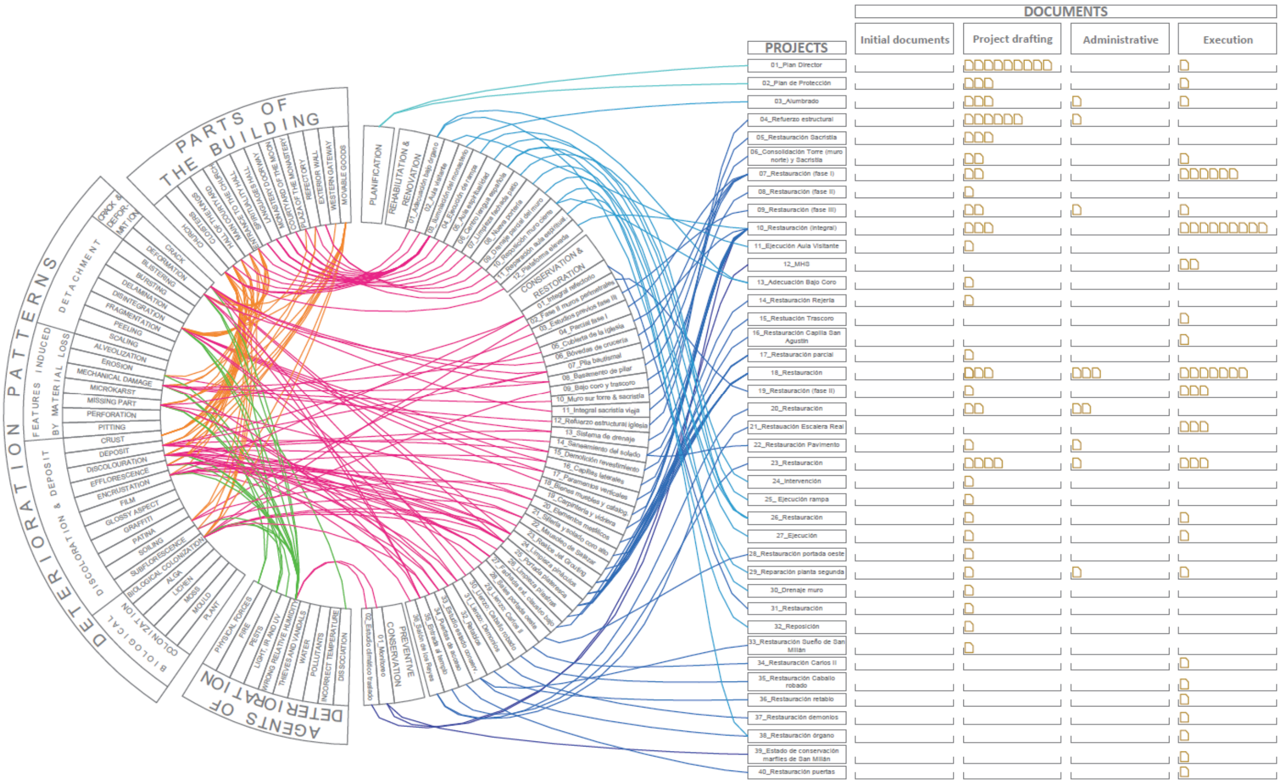

Figure 4 Entities and connections of the graph database represented by means of a radial chart. Colours indicate different types of links: blue for the relationships between conservation-restoration actions and projects, magenta for the relationships between actions and the entities of the physical reality, green for the links of deterioration patterns and decay factors and orange for the relationships between deteriorations and parts of the building

The image of the radial chart is a preliminary version sketched out in a software for vector drawings (CAD). Subsequent steps, once having analysed the communicative interest of the product, will be to generate an interactive version where the connections will be highlighted on demand and that will serve as a friendly interface for searching the database and accessing the documents [27].

4 MATRIX CODIFICATION AND ANALYSIS

Most useful though visual representations are, when there are a large number of vertices and connections, it becomes very difficult to have an overall picture of the complete set of data and, simultaneously, being able to identify local relationships. On another note, graphs are mathematical structures, so they can be also studied using algorithms and statistics that may assist us to extract meaningful information.

During the initial stages of the conceptual model generation, a list of queries that the database should address was compiled. It is advisable that the number of questions is not very long since, this way, it will be easier to define both user interfaces and the own structure of the semantic database. The success lies in avoiding unnecessary complexity and keeping with too generic features which do not add value to the queries [28]. The selected questions for the case of the monastery are listed below (the element underlined is the entry value and the element in bold is the response returned as a result of the query):

1. Where (in which **parts of the monument**) a particular <u>type of deterioration</u> is present?
2. Which **factors of deterioration** are the origin of a <u>type of deterioration</u>?
3. Which **rehabilitation-renovation works** have been done in a <u>part of the building</u>?
4. Which **deteriorations** have been treated in a <u>conservation-restoration work</u>?
5. Which **factors of deterioration** have been controlled in a <u>preventive conservation work</u>?
6. In which **parts of the building** there have been <u>rehabilitation-renovation works</u>?
7. Which is the **project** related to a particular <u>conservation-restoration action</u>?
8. Which are the **documents** related with a <u>project</u>?

From a mathematical perspective, the conceptual model presented in Fig. 1 will tackle these questions by selecting an entity of a particular type (the underlined elements) and searching direct connections (length equal to one) to entities of another type (the ones in bold). For instance, for the first question, a node of the type "deterioration pattern" (e.g., a mechanical damage) is selected and the system returns all the entities of the type "part of the building" that are connected to it.

Normally, the meaning of a direct link may seem obvious and does not need detailed explanations. However, this functionality may seem excessively basic, given the semantic potential of the graph. Consider, by example, the relationships established among the three types of entities within the physical reality. It is interesting to realize that these connections can also be seen as the way to articulate the following questions (Fig. 5): "How?", "Why?", "Where?" and "What?". By associating each connection with its corresponding interrogative term, it becomes possible to formulate more complex queries that integrate multiple elements within the graph. This approach enhances the interpretative capacity of the dataset, allowing for a deeper exploration of the relationships.

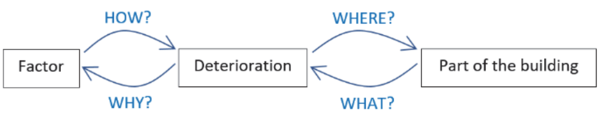

Figure 5 Questions which relate the three types of entities of the "physical reality"

Following the example, in the analysis of a particular deterioration pattern (e.g., a water stain) the question "Where?" directs to the parts of the building where this issue is present, while the question "Why?" informs about the factors that can generate it. Similarly, starting from a factor (e.g., excess humidity) the question "How?" shows the decay patterns induced on the material parts and, finally, beginning with a part of the building, the question "What?" shows the deteriorations that are present.

Once again, these simple questions represent paths of length equal to one and, in essence, they are the links that are directly identified by the visual inspection of the graph. Nevertheless, it is also possible to concatenate links and query about: "How?" + "Where" (i.e., In which parts of the building the deteriorations caused by the factor X are present?), "What?" + "Why?" (i.e., Which factors may be at the origin of the deteriorations that are visible in the part of the building Y?) or, even, questions with more than two components such as: "How?" + "Where?" + "What?" (i.e., Which is the complete list of alterations that appear in the parts of the building where there are alterations originated by the factor X?).

Obviously, the length of the queries can be further extended. In this particular case, this extension allows for an analytical shift beyond the "physical reality" of the building and encompassing the domains of the "actions" and the "documentary archive". However, the difficulty to verbalize the query increases, while the interpretation and meaning of the response becomes more and more complex. Therefore, although the exploration of the graph via long queries is appealing, it is not appropriate to overextend. The optimal choice depends on selecting, among all the possible connections (paths) that we can go through, those that have a particular use, in other words, keeping only with interesting connections [29].

As previously mentioned, graphs are not only suitable for visual representation, but also serve as a valuable tool for analytical processing [30]. The queries may be articulated with specific languages - such as SPARQL, Neo4j, etc., although, on the user's side, it is most useful to have friendly query interfaces that allow interacting with the database using natural language [31, 32].

A suitable way to encode and show the paths of variable length existing in a graph is the adjacency matrix, which is a key tool in the mathematical analysis of graphs [33]. The adjacency matrix (A) is a two-dimensional table in which there are a row and a column for each vertex of the graph, when a (directed) connection from vertex i to vertex j exists, this connection is indicated with a value 1 in the cell situated in the row i column j; the rest of the positions of the matrix are filled in with zeros. The number of paths of a specified length that exist between each pair of vertices is determined by the powers of this matrix.

Nevertheless, the presented conceptual model is not fully adequate for the direct analysis with the adjacency matrix and needs some adjustments. This is due to the role played by the cardinality of the relationships when it comes to discover relevant information that is not evident from a direct inspection. Indeed, there are three different situations that are considered to generate a derived graph from the conceptual model:

a) If the cardinality between two types of entities is from n to n, two-way connections are drawn (i.e., two different question words exist depending on the direction of the course). This happens, for instance, when considering that a kind of "deterioration" can be present in many "parts of the building" and, at the same time, that a "part of the building" can have many patterns of "deterioration".
b) If the cardinality between the elements of the conceptual model is 1 to n, a single connection directed to the n is established. An example is a "project" that is described in several "documents", but a particular "document" is part of only one "project".
c) Finally, in case that the cardinality between the elements of the conceptual model is 1 to 1 both vertices merge. This is what happens with the different types of "interventions" and their specializations as "rehabilitation/renovation", "curative conservation/ restoration" or "preventive conservation".

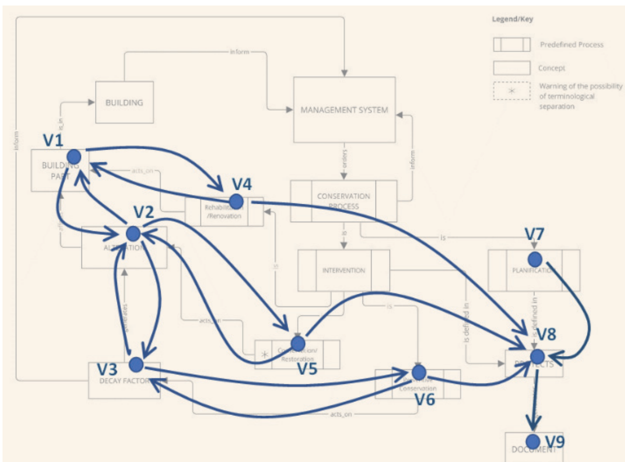

Figure 6 Simplified graph obtained from the conceptual model, used to find paths concerning interesting questions for the management system

With all these considerations in mind, the following directed graph is obtained (Fig. 6) - to simplify the notation, the vertices are identified correlatively (v_1, v_2, v_3…). Over this graph, the possible paths of different lengths will be studied.

Firstly, the adjacency matrix is generated. This matrix indicates the number of paths of length one that can be established from each pair of initial (in the row) and final (in the column) vertices.

$$A = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \quad (1)$$

The square of this matrix (A^2) - this is a matrix multiplication, it is not the square of the values of each entry - shows the number of paths of length two between each pair of vertices.

$$A^2 = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{pmatrix} 2 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 3 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \quad (2)$$

For instance, the A^2 matrix shows that from a vertex v_1 (a "part of the building"), it is possible to reach vertices of the type v_3 ("decay factors", through an intermediate v_2 vertex, i.e., a "deterioration pattern"), vertices of the type v_5 ("preventive conservation works", also via a v_2 intermediate vertex), and vertices of the type v_8 ("projects" that are reached through a node v_4 - an "action of rehabilitation/renovation"). Likewise, there are two different ways to return to v_1 (on the one hand, coming back from a v_2 vertex and, on the other hand, returning from a v_4 node).

Each path can be verbalized using a question like the ones indicated in the paragraph after Fig. 5. Similarly, the cube of the matrix (A^3) shows the paths of length 3 between each pair of vertices and so on.

$$A^3 = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{array} \begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{pmatrix} 0 & 4 & 0 & 2 & 0 & 1 & 0 & 1 & 1 \\ 4 & 0 & 4 & 0 & 3 & 0 & 0 & 2 & 1 \\ 0 & 4 & 0 & 1 & 0 & 2 & 0 & 1 & 1 \\ 2 & 0 & 1 & 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 3 & 0 & 1 & 0 & 1 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \quad (3)$$

The number of possible paths grows very fast as the length increases. However, only the paths regarding relevant questions for the management of the monument - such as the ones indicated in the following picture (Fig. 7) are interesting.

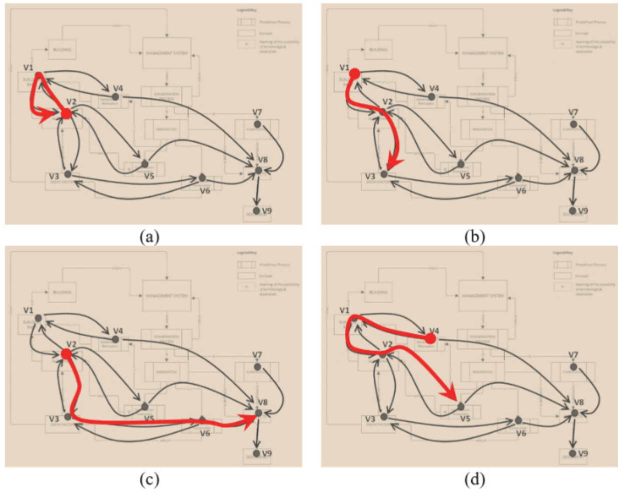

Figure 7 Some examples of paths of length 2 and 3 representing interesting questions for the management of the building: (a) Which other alterations are present in combination with the alteration X (e.g., alveolation)?, (b) Which factors may be active in the part of the building Y (e.g., the cloister)?, (c) Which projects of preventive conservation were carried out in order to mitigate the factors which can trigger the alteration Z?, (d) Which actions of palliative conservation have acted on alterations that are active in the area where the action of rehabilitation W was carried out?

Nonetheless, there will be questions that the structure will be unable to process since the way the connections were established makes it impossible to complete the search o because the questions do not correspond to the real meaning of the existing links (Fig. 8).

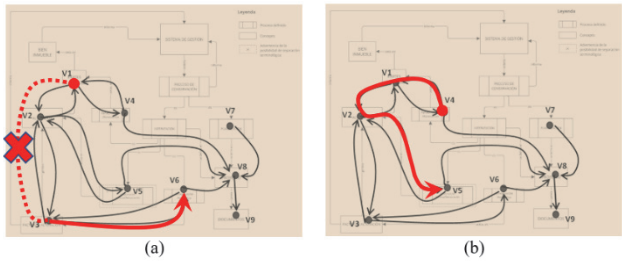

Figure 8 Examples of questions that the graph is not able to answer correctly: (a) Which preventive conservation works have been carried out in a particular part of the building (e.g., the church)?, the issue here is that there is not a direct relationship between "parts of the building" (v_1) and "factors" (v_3) which are the elements on which the preventive conservation is applied, so the question cannot be formed without the reference to an intermediate "alteration" (v_2) that is presented in this part of the building, (b) Which works of palliative conservation are directly linked to a particular restoration work?, in this case, the cardinality of the intermediate connections makes it impossible to determine whether the works of palliative conservation (v_5) retrieved by the system are or not directly linked with the original node concerning a restoration work (v_4), note that the previous figure included a different question for this very path depicted over the conceptual model, that one is adequate for the relationships that can be solved

5 DISCUSSION

Two complementary ways of visualizing a knowledge graph have been presented: a network and a radial layout. The former, network display, permits focusing on the way that the entities are grouped according to the number of common links, while the latter, radial layout, fixes the position of the nodes beforehand (allowing, this way, reflecting a particular classification of the entities). The network display is closely related with other structures such as classification charts and trees; hence, networks are easily adapted to visualization and searching tools that make use of these structures; likewise, as the positions of the nodes are automatically assigned, they are immediately redrawn when the database is updated (for instance, when new nodes and links are added). However, what is better represented in a network are the groups, but it takes a long time to locate where the different elements are placed, for that reason, the radial display is preferable when the users are looking for particular entities.

Concerning the formulation of the queries from the analysis of the possible paths that can be traced thanks to the adjacency matrix, the previous examples reveal that the design of a conceptual model entails decision-making on the meaning of the relationships; in turn, the meanings condition the questions that can be posed. Certainly, the analysis of the possible questions related to the paths in the graph allows checking whether the conceptual model is suitable for the needs of the expected management of the heritage element.

Likewise, the fast increase in the number of paths with the considered length should always be taken into account, as well as the difficulty of formulating the corresponding questions. Therefore, simplified versions of the graphs may be helpful for the studies.

Most previous references on conceptual modelling have focused on knowledge formalization, i.e., on gathering experts' perspectives to identify significant entities and relationships; however, limited attention has been given to the manner in which these connections are established. This study proposes a methodology to address this gap, which consists of the following steps:

1. Firstly, the conceptual model is transformed into a processable graph considering the cardinality of the links. In parallel, each link is associated with a question word.
2. The number of connections between each pair of nodes is computed with the adjacency matrix and its powers. Then, the connections are verbalized using the question words of the successive stretch of the path.
3. From all possible paths, only interesting connections are retained.
4. A redesign of the conceptual model is suggested when necessary connections are missing, or in case that the verbalization of the paths disagrees with the expected meaning.
5. Once the conceptual model is fixed and the database populated with the specific elements of the study, it is time to tackle the visual representation, for which many valuable options are available.

In the particular case of the monastery of San Millán, the final aim is to integrate the visualizations into a computer tool that will be accessible to the institution responsible for the management of its conservation-restoration. Thus, there is a need to conjugate the design of the database and the real possibilities of implementation, considering the resources for the development, user training (organization staff) and long-term maintenance, among other aspects.

6 CONCLUSIONS

When representing a knowledge graph through static or interactive visualisations, a wide range of graphical configurations can be generated, depending on the spatial arrangement of its elements and the additional information conveyed through visual variables such as colour, size, value, orientation, shape, and texture. Therefore, the investment of time for the design and the customization of the representation, rather than relying on default layouts generated by the software available, proves to be highly beneficial. Furthermore, multiple complementary representations emphasizing specific characteristics and groups of the database are possible, which can be interesting to enable different management tasks. This strategy allows more detail and efficient depiction of the knowledge contained within the graph.

Usually, graphs are considered as pure topological structures, in which the position of the elements is irrelevant since all the information is vested in the connections, this is the situation represented by the initial network display. On the contrary, sometimes, the placement of the elements provides meaningful information and facilitates the comprehension of the complete set; this is what is shown with the radial display.

Moreover, it is noteworthy that certain features of the graph such as the existence of partitions of the nodes or the cardinality of the relationships in the conceptual model are important both for the selection of the type of visualization and for the mathematical analysis. Therefore, their study is essential to achieve a cohesive visual representation and analytical processing.

In conclusion, graph databases represent a rapidly evolving field undergoing advancements in semantics, mathematical, and communicative aspects. Nevertheless, future research should continue exploring ways to integrate these dimensions comprehensively. After all, the key finding is that cases as the one presented in this text (information management of a cultural heritage element) evince that users approach the graph to analyse the knowledge about a particular subject (here, the "semantical" aspect of the graph is presented), then data processing requires computing (i.e., resorting to the "mathematical" aspect), and, finally, the results need to be presented in an understandable and user-friendly manner (thus, the "communicative" aspect must also be involved).

Acknowledgements

The authors acknowledge the support of the Fundación San Millán de la Cogolla for the realization of this study. The participation of JK in this research is supported by the Basque Government through a grant for doctoral studies in the year 2019-2020 and successive renewals.

7 REFERENCES

[1] Kerren, A., Purchase, H. C., & Ward, M. O. (2014) Introduction to multivariate network visualization. *Multivariate Network Visualization*, *8380*, 1-12. https://doi.org/10.1007/978-3-319-06793-3

[2] Robinson, I., Webber, J., & Aifrem, E. (2015). *Graph databases. New oportunities for connected data*. O'Reilly.

[3] Korro, J., Valle-Melón, J. M., & Rodríguez-Miranda, Á. (2024). Documentary data collection: an initial step for information management in the conservation and restoration of cultural heritage. *Conservar Patrimonio*. https://doi.org/10.14568/cp27370

[4] Cacciotti, R., Blaško, M., & Valach, J. (2015). A diagnostic ontological model for damages to historical constructions. *Journal of Cultural Heritage*, *16*(1), 40-48. https://doi.org/10.1016/j.culher.2014.02.002

[5] Fiorani, D. & Acierno, M. (2017). Conservation process model (CPM): A twofold scientific research scope in the

information modelling for cultural heritage. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *42*(5W1), 283-290. https://doi.org/10.5194/isprs-Archives-XLII-5-W1-283-2017

[6] Messaoudi, T., Véron, P., Halin, G., & De Luca, L. (2018). An ontological model for the reality-based 3D annotation of heritage building conservation state. *Journal of Cultural Heritage*, *29*, 100-112. https://doi.org/10.1016/j.culher.2017.05.017

[7] Colucci, E., Kokla, M., & Noardo, F. (2021). Ontology-based data mapping to support planning in historical urban centres. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B4-2021), 39-45. https://doi.org/10.5194/isprs-archives-XLIII-B4-2021-39-2021

[8] Fiorani, D., Acierno, M., Cutarelli, S., & Donatelli, A. (2021). Transformation of tools and conservation of architecture: some researches on the use of digital systems for the intervention on the historical buildings. *Építés - Építészettudomány*, *49*(1-2), 97-131. https://doi.org/10.1556/096.2020.005

[9] ISO 21127 (2023). Information and documentation. A reference ontology for the interchange of cultural heritage information.

[10] Moraitou, E. & Christodoulou, Y. (2021). *Linked Conservation Data: LCD Modelling Working Group Overview of current conservation and restoration models* (Issue March).

[11] Faraj, G. & Micsik, A. (2021). Representing and Validating Cultural Heritage Knowledge Graphs in CIDOC-CRM Ontology. *Future Internet*, *13*, 277. https://doi.org/10.3390/fi13110277

[12] Bai, N., Nourian, P., Luo, R., & Pereira Roders, A. (2022). Heri-Graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. *ISPRS International Journal of Geo-Information*, *11*, 469. https://doi.org/10.3390/ijgi11090469

[13] Jakun-Kelly, T. J., Dwyer, T., Holten, D., Hurter, C., Möllenburg, M., Weaver, C., & Xu, K. (2014). Scalability considerations for multivariate graph visualization. *Multivariate Network Visualization*, *8380*, 207-235. https://doi.org/10.1007/978-3-319-06793-3

[14] Wybrow, M., Elmqvist, N., Fekete, J.-D., Von Landesberger, T., van Wijk, J. J., & Zimmer, B. (2014). Interaction in the visualization of multivariate networks. *Multivariate Network Visualization*, *8380*, 97-125. https://doi.org/10.1007/978-3-319-06793-3

[15] Shi, L., Liao, Q., Tong, H., Hu, Y., Wang, C., Lin, C., & Qian, W. (2020). Onion Graph: hierarchical topology + attribute multivariate network visualization. *Visual Informatics*, *4*, 43-57. https://doi.org/10.1016/j.visinf.2020.01.002

[16] Windhager, F., Federico, P., Schreder, G., Glinka, P., Dörk, M., Miksch, S., & Mayr, E. (2019). Visualization of cultural heritage collection data: state of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics*, *25*(6). https://doi.org/10.1109/TVCG.2018.2830759

[17] Yoghourdjian, V., Archambault, D., Diehl, S., Dwyer, T., Klein, K., Purchase, H. C., & Wu, H. Y. (2018). Exploring the limits of complexity: a survey of empirical studies on graph visualization. *Visual Informatics*, *2*, 264-282. https://doi.org/10.1016/j.visinf.2018.12.006

[18] Roberts, J. C., Yang, J., Kohlbacher, O., Ward, M. O., & Zhou, M. X. (2014) Novel visual metaphors for multivariate networks. *Multivariate Network Visualization*, *8380*, 127-150. https://doi.org/10.1007/978-3-319-06793-3

[19] McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., & Pinaud, B. (2019) State of the art in multilayer network visualization. *Computer Graphics forum*, *38*(6), 125-149. https://doi.org/10.1111/cgf.13610

[20] Muelder, C., Gou, L., Ma, K. L., & Zhou, M. X. (2014). Multivariate social network visual analytics. *Multivariate Network Visualization*, *8380*, 37-59. https://doi.org/10.1007/978-3-319-06793-3

[21] Bertin, J. (1967). *Semiologie graphique*. Editions Mounton & Cie.

[22] Robinson, A., Morrison, J., Muehrcke, P., Kimerling, J., & Guptill, S. (1995). *Elements of cartography*. John Willey & Sons Inc.

[23] Adrienko, N. & Adrienko, G. (2006). *Exploratory analysis of spatial and temporal data. A systematic approach*. Springer.

[24] Pretorious, A. J., Purchase, H. C., & Stasko, J. T. (2014) Task for multivariate network analysis. *Multivariate Network Visualization*, *8380*, 77-95. https://doi.org/10.1007/978-3-319-06793-3

[25] ICOMOS & ISCS (2008). *Illustrated glossary on stone deterioration patterns*.

[26] IIC-ITCC: The Canadian Conservation Institute. (2018). *Preventive conservation: agents of deterioration*, *2*(1).

[27] Ávila-García, L., Ortiz-Repiso, V., & Rodríguez-Mateos, D. (2015). Herramientas de descubrimiento: ¿una ventanilla única? *Revista Española de Documentación Científica*, 38(1): e077. doi: http://dx.doi.org/10.3989/redc.2015.1.1178

[28] Debruyne, C., Kilgallon, L., O'Sullivan, D., & Crooks, P. (2022). Creating a knowledge graph for Ireland's lost history: knowledge engineering and curation in the Beyond 2022 project. *Journal on Computing and Cultural Heritage*, *15*(2), 25. https://doi.org/10.1145/3474829

[29] Hyvönen, E. & Rantala, H. (2019). Knowledge-based relation discovery in cultural heritage knowledge graphs. *Digital Humanities in Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 4th Conference, Copenhagen, Denmark*, *2364*, 230-239.

[30] Bruschke, J. & Wacker, M. (2014). Application of a Graph Database and Graphical User Interface for CIDOC CRM. *CIDOC 2014 annual conference*.

[31] Pellegrino, M.A., Scarano, V., & Spagnuolo, C. (2023). Move cultural heritage knowledge graphs in everyone's pocket. *Semantic Web*, *14*, 323-359. https://doi.org/10.3233/SW-223117

[32] Fafalios, P., Kritsotaki, A., & Doerr, M. (2023). The SeaLiT Ontology - An Extension of CIDOC-CRM for the Modeling and Integration of Maritime History Information. *Journal on Computing and Cultural Heritage*. https://doi.org/10.1145/3586080

[33] Goodaire, E. G. & Parmenter, M. M. (2006). *Discrete mathematics with graph theory*. Pearson.

Contact information:

Álvaro RODRÍGUEZ-MIRANDA, Engineer in Geodesy and Cartography, Doctor and Adjunct Professor
(Corresponding author)
Experimental space for innovation in Heritage and Cultural Landscapes (GPAC),
Department of Applied Mathematics,
University of the Basque Country (UPV/EHU),
Justo Velez de Elorriaga, 1 (Micaela Portilla building),
01006 Vitoria-Gasteiz, Spain
E-mail: alvaro.rodriguezm@ehu.eus

Jaione KORRO, Conservator and Restorer of Cultural Heritage, Doctor and Assistant Lecturer
Experimental space for innovation in Heritage and Cultural Landscapes (GPAC),
Department of Painting,
University of the Basque Country (UPV/EHU),
Justo Velez de Elorriaga, 1 (Micaela Portilla building),
01006 Vitoria-Gasteiz, Spain
E-mail: mirenjaione.korro@ehu.eus