Semantic Network Analysis as a Knowledge Representation and Retrieval Approach Applied to Unstructured Documents of Construction Projects

R.R. Aragao[a], and T.E. El-Diraby[b]

[a]Petrobras – Petróleo Brasileiro S.A., Brazil; University of Toronto, Canada (alumnum)
[b]Department of Civil & Mineral Engineering, University of Toronto, Canada
E-mail: r.aragao@petrobras.com.br, r.aragao@alum.utoronto.ca, tamer@ecf.utoronto.ca

Abstract –

Reusing knowledge from past projects is a critical task in construction, given the increasing complexity in such projects: numerous stakeholders, a multi-disciplinary domain, and multi-objectives besides the traditional ones such as cost and schedule. Unstructured data, such as progress reports and minutes, is a rich source of knowledge that can be revisited in projects as the contextual nature of documents permits describing the nuances of the interrelations and uniqueness of each project. However, texts are difficult to formalize in a way that the process of retrieval and analysis of relevant knowledge be automated using computers. In this paper, we present an innovative approach that encompasses formalization, retrieval, analysis, and reuse of knowledge from case studies of past construction projects. Assuming that energy is an objective, the cases are represented as a network of common concepts found in every project. The nodes are the concepts, and the links between them are established whenever there is an association between two concepts that affects the energy use in construction. Conversely, the comments of team members of a current project can also be captured and represented using the same standardized set of concepts. Using network analysis, we can retrieve the most relevant cases, which are similar to the current project, study the most important concepts, extract clusters of concepts, and capture the nuances of the cases in a more objective way. A concept map based on the literature and three case studies of past oil and gas projects are developed to undertake this approach. We evaluate the method by simulating the collaborative environment of one of the cases through the participation of ten volunteers in Green 2.0, an online media to discuss construction projects. At the end of the test, we perform a correlation between the networks of the test and the case study.

Keywords –

Knowledge retrieval and representation; Unstructured data; Network analysis; Blockmodeling; Construction management; Energy use

1 Introduction

The primary objective of this work is to utilize network theory to help formalize and consistently process unstructured data (mainly text) in construction management. The aim is to support capturing, retrieving and coordinating knowledge reuse during project deliberations. Documents and reports developed based on the input or deliberations of project stakeholders are a fundamental container and widely used means to capture construction knowledge. This fact is primarily due to the subjectivity of the domain. In general, planning and managing construction projects depend on the reuse of implicit knowledge and expertise gained in previous projects. The challenge is how we can capture and reuse knowledge contained in unstructured data developed by project stakeholders. How can one externalize the tacit knowledge contained in them? Doing so in an efficient manner has been increasingly becoming important due to the growing complexity of projects. First, due to the increase in the number of stakeholders who have to participate in decision making. Second, the increase in the number and diversity of decision criteria. For example, in addition to the traditional cost and schedule, project decision makers have now to consider additional issues such as environmental impacts as well as energy consumption. Third, compounding the challenges, many of the issues are contested—for example: how to identify and measure impacts on communities and sustainability.

Foundational approaches, such as rule-based expert systems and knowledge bases, have significant limitations in this milieu [1]. These limitations are due to the contextual nature of projects. Each project has unique composition/attributes and boundary conditions, which influence the relevance and interpretation of rules as well as the ability to adapt them to the needs of the project context. Semantic models of knowledge, such as ontologies, provide a very suitable approach to handle such subjective and unstructured nature of knowledge [2]. However, the static nature of ontology has shortcomings [3]. This constraint, again, is due to the contextual nature of construction knowledge. More importantly, knowledge constructs (models) are "relativistic" as they depend on the judgment of stakeholders: it is typically quite difficult to achieve full (standardized) agreements between stakeholders.

Unstructured data, such as online discussions, minutes of meetings, e-mails, progress reports, and case studies have recently gained the attention of knowledge management (KM) practitioners [4,5]. This remark is because such corpus represents subjective knowledge in semantic forms. They are also, by default, context-sensitive. However, possibly, the most important advantage of using project text corpus as a source of knowledge is that they are the results of intense and critical dialogue between experts. Meetings transcripts, online discussions, e-mail threads, final reports and project case studies represent venues for knowledge exchange, conflict resolution, innovation, and consensus creation. Hence, it is not surprising that unstructured data are more informative than other sources.

Given that a good deal of the knowledge in the domain is textual, can using semantic network analysis be helpful? Instead of free text, typically used in case studies, can we build networks of project concepts and benefit from the traditional network analysis measures in enhancing the formality and objectivity of analyzing project knowledge? For example, what can we learn about a project if we study the centrality of, say, safety in its concept network? If two concepts, for instance, safety and excavation, are not linked in the concept network, does this reflect incomplete analysis?

In other words, the specific problem that this research project aims to address is that pure textual representation of project knowledge in the case studies may not be an optimal solution. First, the retrieval of such cases for future use can be inefficient. Typically, this would be done through a set of keywords recorded by the case developers or suggested by a content analysis algorithm. The automatically generated topics are suggested based on word frequencies and associations. This procedure, however, does not necessarily capture coherent or essential topics from a text corpus. While human tags (the manual categorization of the text in meaningful topics) are more superior, their standalone nature does not allow for capturing the interrelationships of case concepts. Furthermore, such tags are never optimal (they cannot cover all possible future needs). Second, it is difficult for upper management to conduct any formalized analytics using just a set of documents. Beyond the benefits of case reuse to guide new projects, cross-case analysis can detect trends, suggest and emphasize concept relationships (for a more detailed review of works published in this field, please see [6]).

Compared to other stages of the project life cycle, the construction phase has received the least attention from researchers [7]. This fact is mainly because of its limited contribution to the overall energy consumption of a project but, more importantly, it is the unique challenges of the construction stage. While the energy use in construction has been a topic of constant debate in academia [8,9], this scope falls much more under the umbrella of the construction knowledge management domain than energy efficiency or life cycle assessment. Hence, this research is not an energy assessment or a quantitative estimate of the energy use of the construction phase.

For such a subjective domain, learning from historical project cases can be very helpful. They map prior project conditions and their relationships to project performance (about energy); and document pitfalls, solutions, and lessons learned. The textual nature of these cases allows the user to contextualize the options and challenges of the project and gain a deeper understanding of the relevance of the listed best practices to a new project.

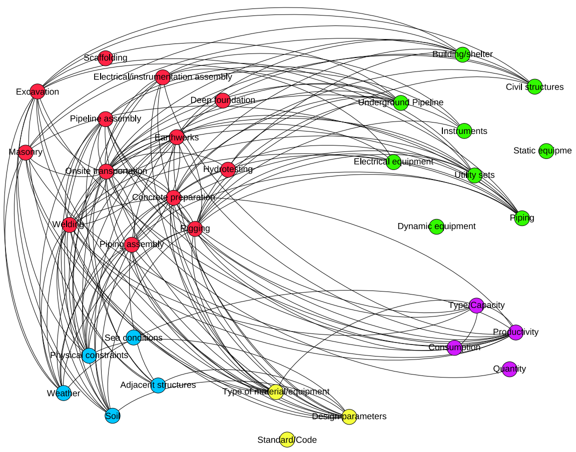

Figure 1. The concept network for one of the case studies developed in the research

The main objective of this research is to realize a system to use network theory to help formalize text corpus of project cases and, possibly, other documents to facilitate the externalization of tacit knowledge in an objective and reusable manner. The main contribution of this research is the creation of an approach that formalizes the knowledge of projects embedded in unstructured data through transferring text into a network of concepts. In more specific terms, by using the formalized networks of concepts, the method can capture and reuse the knowledge from previous project cases to support the deliberations of the decision-making team during the construction planning of future projects.

The proposed system will permit stakeholders to collaborate, investigate, formalize, and capture the knowledge embedded in unstructured data to manage energy use in the construction phase of O&G projects.

2 Methodology

The proposed approach relies on the following scenario: using manual or automated means, the network of concepts of a current project is established. The network can be developed through semantic analysis of project documents or a summary report of a project, hopefully, developed by a group of collaborators. At the same time, during a retrieval process, the project team was able to find relevant cases that are similar to the current project. Figure 1 shows an example of the graph of a network of concepts of one of these case studies. The team then is interested in discovering (or externalizing) the most relevant knowledge constructs of this case. The goal is to help them sift through the complexity of the network: what are the concepts that have significant interaction within the current project. Alternatively, help them externalize and discover some of their implicit knowledge that the network captured for them. For that, they use a prior generic network built on the common sense knowledge of a group of experts. Blockmodeling, in this case, is used to rearrange this generic network into meaningful blocks that are easier to visualize and to interpret. A manager can work on comparing the blocks of several cases, the current project, and the generic network to study differences of patterns.

The main steps of the methodology are as follows:

a) Based on the literature review, we present a concept map, which is a simplified taxonomy of concepts associated with the energy use in construction. The proposed concept map guarantees the networks are always represented on the same base of concepts Figure 2.

b) We develop three case studies of past oil and gas projects in Brazil. In-depth semi-structured interviews of project team members are the main source of evidence of each case. A minimum number of participants should be interviewed until the authors reach the saturation of information, in which the information provided by the interviewees begin to repeat and therefore no additional interview is necessary [10]. The participants respond open-ended questions about the challenges, best practices, opportunities, and risks that affects or could affect the energy use during the construction phase. A profile of responses is generated for each interview. In each profile, the research team identifies keywords or expressions that semantically resembles the concepts from the base taxonomy (letter “a”). A relation or a link is established whenever there is a semantic association between two concepts that affects the energy use in the construction phase. These relations are transferred to an adjacency matrix, and a network of concepts is represented for each case study, such as the one in Figure 1. The procedure above permits to represent any case study as a network of concepts. Dealing with case studies as a formalized network of pre-established concepts allows us to apply several metrics and indexes to study them by using network analysis. For example, if the discussions or social interactions of members of a current project, such as an online discussion tool, are captured using the same procedure discussed in letter “b”, a QAP analysis [11] of the current and case networks offers an interesting retrieving method of similar past projects.

Supposing that the three case networks presented in this article were retrieved from a repository of cases using such alternative, now we want to compare the retrieved cases with a generic network, which does not reflect any specific project. Exploring the differences between project-driven cases and a general network can shed light on what relevant knowledge the past project has to offer so that it should be taken into consideration in the new project.

c) To collect the data for the generic network, we conduct an expert survey [6]. As opposed to the project-related cases, the survey is conducted to collect the common knowledge of the participants without any project in mind. To do so, we transform the 573 possible relations of the concept map in close-ended questions, divided them into fourteen online surveys and submit them to experts in the O&G field in Brazil. The questions are multiple choice, in which the answers follow an intensity rating scale: no effect/low effect/moderate effect/high effect, with each answer having a score from 1 to 4, respectively. As a criterion to create the relations in the generic network, we establish a link or tie between two concepts whenever the average of all responses for each question was greater than 2.5. Three examples of questions are listed below.

- Does excavation affect the energy use during the construction of utility systems?
- Does welding affect the energy use during the assembly of piping?
- Does weather affect the energy use of earthworks?

d) Once the data collection phase is concluded, we represent the survey answers as a three-mode generic network, in which the ties between the nodes of the same category are disregarded.

e) The steps below allow us to contrast the knowledge constructs between each case and the generic network, by using a three-mode generalized blockmodeling approach [12] to create a baseline (a blockmodel) from the generic network as follows:

- We pre-define the clusters of the construction activities based on their typical function in construction sites and our prior knowledge of the cases. This assumption increases the chance to find substantive blocks using the blockmodeling method.
- As an input for the blockmodeling, we define the number of clusters and the type of blocks allowed to form during the optimization. With the clusters of the activities and the constraints established beforehand, the algorithm attempts to reorder the remaining concepts in such a way that it fulfills the defined requirements.
- After running the blockmodeling with the constraints defined above for the generic network, we select the solution that provides the best interpretation. The status of generic network is assigned to the best solution.
- The adjacency matrix of the case networks are rearranged to fit the blockmodel obtained from the previous procedure. In this case, the interest is in the different blocks. The different blocks and their inter-relations are interpreted and compared with the baseline in the light of the case studies, and then these findings are reported.
- We introduce a block analogy index that quantitatively assesses the overall and the block-to-block similarities of the case-based networks. The less similar the case network is to the baseline, the more case-specific, contextual the project is.

Construction Activities		Systems		Factors	
1	Excavation	14	Civil structures	Design	
2	Deep foundation	15	Building/shelter	23	Type of material/equipment
3	Welding	16	Static equipment	24	Design parameters
4	Piping assembly	17	Underground pipeline	25	Standard/Code
5	Earthworks	18	Piping	Site characteristics	
6	Rigging	19	Electrical equipment	26	Soil
7	Concrete preparation	20	Instruments	27	Sea conditions
8	Electrical/instrumentation	21	Dynamic equipment	28	Weather
9	Pipeline assembly	22	Utility sets	29	Adjacent structures
10	Hydrotesting			30	Geographical location
11	Masonry			Resources	
12	Onsite transportation			31	Type/Capacity
13	Scaffolding			32	Consumption
				33	Productivity
				34	Quantity

Figure 2. The proposed concept map (taxonomy) of energy concepts

3 Results and Discussions

The results are presented in the following sections for the concept networks of projects A, B, and C, and the combined case. More details of the case studies can be found in [6]. The combined case is a Boolean sum of the three networks of cases, and is an attempt to produce a network that generically represents the three cases. Three analyses are presented: similarity, centrality-level, and blockmodeling. Detailed information for each analysis can also be found in [6].

3.1 Similarity of Networks

We used the concept of dissimilarity to develop a retrieving process. Two nodes have a dissimilarity equal to 1 (one) if they do not share any one of their neighbors. In contrast, they are structurally equivalent (or 100% similar) if the dissimilarity is 0 (zero) or, in other words, they share all their ties.

Using UCINET, we used the Quadratic Assignment Procedure (QAP) regression analysis to calculate the correlation coefficient and its respective p-value for each pair of networks. In the dissimilarity matrices, each cell of the matrix is the dissimilarity calculated for each pair of nodes of the original studied matrix). The dissimilarity is zero for the elements of the diagonal of the matrix, since each node is 100% similar to itself.

Table 1. Correlation coefficient of the dissimilarity matrices using QAP regression of the three case studies and two combined case networks

	Project A	Project B	Project C	Combined case 1	Combined case 2
Project A	1	0.516	0.309	0.720	0.695
Project B	0.516	1	0.333	0.690	0.728
Project C	0.309	0.333	1	0.502	0.433
Combined case 1	0.720	0.690	0.502	1	0.659
Combined case 2	0.695	0.728	0.433	0.659	1

Table 2 P-values of the similarity matrices using QAP regression

	Project A	Project B	Project C	Combined case 1	Combined case 2
Project A	0	0.0003	0.0003	0.0003	0.0017
Project B	0.0003	0	0.0003	0.0003	0.1310
Project C	0.0003	0.0003	0	0.0003	0.1060
Combined case 1	0.0003	0.0003	0.0003	0	0.0017
Combined case 2	0.0017	0.1310	0.1060	0.0017	0

Table 1. presents the correlation coefficient for each pair of the similarity matrices using the QAP regression

on UCINET. Each concept network has a similarity matrix. Since the correlation coefficient can assume any value between -1 and 1, the values in the diagonal are one because, indeed, the correlation between each matrix and itself is perfect. Nevertheless, the results of Table 1 must be carefully cross-checked with the p-values in Table 2. To be confident that a strong correlation exists between the structure of two networks, assuming a 95% confidence interval, the p-values in should be lower than 5% or 0.05, meaning that such correlation coefficient is statistically significant.

By analyzing the results, it is observed that the network of Project A correlates better with Project B when compared to Project C (0.516 against 0.309). This correlation probably exists because, although Project A and B have different logistics and environments, both construction projects are characterized by having a high volume of piping, static equipment, electrical/automation, and utility works. The correlations between the case studies are not so little that one can affirm the networks are not part of the same domain, but they are not high enough to be considered useless in dealing with different cases whose network representations are the same.

Regarding the combined cases, each one being an attempt to represent the three case studies generically, the results of Table 1 for the correlation coefficient were considered relatively satisfactory with some limitations [6].

3.2 Centrality Measures

Centrality measures are used to identify the most prominent nodes in the network. The structure of the concept networks suggests that the choice of the most suitable measure depends on the category of node. As the concept networks are directed, we opted to calculate the out-degree, which just considers the number of outward ties. The construction activities are in the intermediate layer of the network, and they serve as a bridge connecting the factors and the physical system; thus, betweenness is more appropriate to detect the more important construction activities. Regarding the systems, we considered not only the nodes with the highest degree, but also the degree of the nodes each node is connected to. In this sense, the in-eigenvector centrality is more indicated [10].

Table 3 presents the out-degree, betweenness, and in-eigenvector centralities for Project A. The table is truncated for the sake of saving space. Concerning the 13 construction activities, rigging (or material handling), onsite transportation, and scaffolding are the most central activities considering the betweenness centrality. These activities were cited several times by the respondents during the interview phase. However, no participant directly indicated the importance of these construction activities as being the most influential for the energy expenditure during construction. This fact could be better investigate by looking at their betweenness centrality. As a brownfield located in an existing industrial plant in full operation, the performance of this project was highly influenced by the numerous restrictions regarding access control, limited layout, and existing equipment or buried structures. These limitations ended up shaping the logistics of the construction, which is formed by the transportation of materials and workforce, material handling, and assembly/disassembly of scaffolds due to work at higher heights. Using the betweenness centrality permitted the research to elicit other influential activities other than the obvious energy-intensive ones, such as welding, excavation, deep foundation, and more.

Table 3. Out-degree, in-degree, betweenness, and in-eigenvector centralities for the network of Project A. The values are presented in descending order to highlight the most central activities/factors

Activities/factors	Out-degree	Betweenness	In-Eigenvector
Construction activities			
Rigging	9	73.261	0.363
Onsite transportation	8	28.152	0.246
Scaffolding	7	19.894	0.297
Welding	4	8.753	0.363
Masonry	9	6.673	0.066
Excavation	8	6.651	0.066
Physical systems			
Piping	0	0.000	1.000
Static equipment	0	0.000	0.964
Civil structures	0	0.000	0.938
Utility sets	0	0.000	0.860
Electrical equipment	0	0.000	0.654
Instruments	0	0.000	0.654
Design factors			
Design parameters	12	16.012	0.000
Type of material/equipment	9	0.000	0.000
Standard/Code	5	0.000	0.000
Site characteristics			
Geographical location	15	0.000	0.000
Adjacent structures	13	0.000	0.000
Soil	5	0.000	0.000
Resources factors			
Productivity	11	49.012	0.117

Type/Capacity	6	51.975	0.206
Consumption	6	4.000	0.117
Quantity	1	0.000	0.000

The out-degree of site characteristics, design and resources factors is the most appropriate centrality to highlight the most prominent factors. The schedule of this project was highly challenging, and the volume of activities provided by the design (scope) had a critical role: the larger the scope is, the higher the energy use. As such, the centrality analysis revealed that the node "design parameters" has the highest out-degree in the design category. Regarding site characteristics, the geographical location and adjacency structures have the highest out-degree. These two factors are closely related to the causes that affected the construction activities mentioned above (access control, limited layout, and the presence of buried structures).

Looking at the in-eigenvector measures of the physical systems, it was found that piping, static equipment, and civil structures were the most impacted systems concerning energy use. As a matter of fact, these four systems encompassed most of the scope of the project. In this regard, static equipment is a proxy for the five atmospheric storage tanks, which required large volumes of rigging, onsite transportation, scaffolding, and welding. Piping and utility systems are portrayed here because of the amount of piping assembly, which also involves welding and energy-intensive logistics. Finally, civil structures received a high in-eigenvector degree because of the works regarding excavation and deep foundation services for the tanks and other pieces of equipment.

3.3 Blockmodeling

To find a meaningful blockmodel for the generic network that can be compared with the networks of the case studies, four subgroups of activities are created according to a functional classification: logistics, heavy-duty equipment, structural civil works, and electro-mechanical assembly. Onsite transportation, rigging, and scaffolding are the three activities that best represent the logistics in construction sites. Excavation, earthworks, and pipeline assembly are characterized by demanding pieces of heavy-duty equipment, such as excavators, dozers, graders, and pipelayers. Structural civil works encompass deep foundation, concrete preparation, and masonry works. Finally, electro-mechanical assembly, which significantly differs oil and gas projects from other installations, is comprised of electrical/instrument assembly, piping, welding, and hydrotesting. In the optimization blockmodeling process, these four subcategories of activities are four pre-established clusters. Hence, the blockmodeling algorithm attempts to create blocks by permuting systems and factors to match the pre-specified clusters of activities, the required of blocks types and the constraints. Since the logistics, heavy-duty equipment, structural civil works, and electro-mechanical services have different characteristics, it is expected that the final optimized blocks capture clusters of meaningful knowledge constructs that are representative of the generic network.

Figure 3 brings the adjacency matrix of the generic network before and after the generalized blockmodeling. We can visualize the participants' perception with regards to the region of the factors vs. activities blocks (area "1"), as well as the "activities x systems" blocks (area "2"). Most importantly, through the figure, we observe how blockmodeling rearranged the nodes in a way that it is easier to interpret and simplify the network.

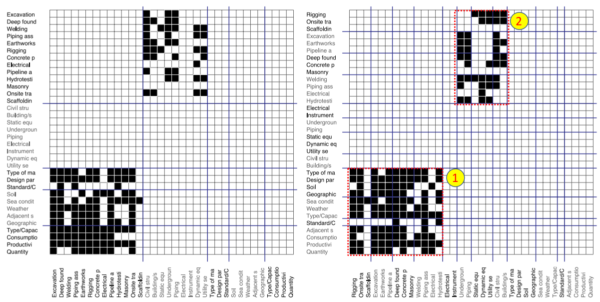

Figure 3. The three-mode adjacency matrix of the generic network before (left) and after (right) a generalized blockmodeling with twelve clusters [13]

Figure 4 (see [13] for enlarged pictures) brings the blockmodel in the form of the adjacency matrix for both the generic network of the survey and Project B. The results suggest Projects A and B have several features in common because many of their blocks diverge equally from the generic network. They have similarities even though the former is an industrial facility and the latter is an offshore LNG terminal.

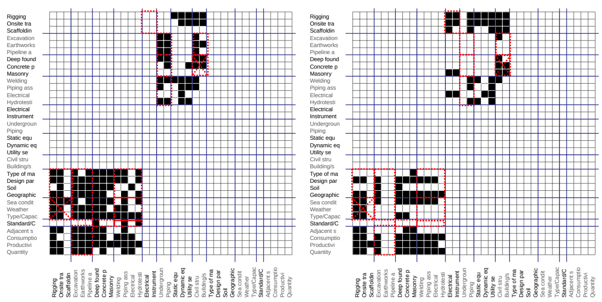

Figure 4. The three-mode adjacency matrix of the blockmodel of the baseline (left) and the rearranged Project B network (right) [13]

The main objective of the survey was to capture the general knowledge of the participants with no specific project in mind, which means that one should not expect the generic network to represent unusual relations and blocks that can potentially aggregate a meaningful knowledge. Conversely, comparing the differences between blocks can pinpoint what is so singular in the case study, and this information, along with the related context of the project, can be used in future projects.

The proposed method also includes the calculation of the average block analogy index $\bar{k}$ for each case-based network with the blockmodel obtained above as a baseline [6,13]. The smaller $\bar{k}$, the more the case-based network moves away from the baseline blockmodels, and therefore the more one can learn from the differences and the context of the past project embedded in a case network.

3.4 The Evaluation Approach

Since we introduced a new collaborative approach for energy-related processes in the construction phase, the best way to assess this research is through evaluation, not validation. Validation checks if a proposed model complies with internal specifications and if the results collected from the samples can be statistically generalized to the population. The validation would require a significant number of projects and a corresponding extensive database to conclude, which is out of the scope of this research.

To undertake the evaluation phase of this research, we conducted a test using Green 2.0, an online discussion media to discuss energy in construction [14]. We simulated the discussion environment of one of the cases studied in this research: Project A. Ten volunteers participated in the test: undergraduate and master's civil engineering students, construction managers, and a design engineer. The participants were educated in the case and asked to tag their comments based on the 34 concepts of the concept map. For each comment, the participants needed to choose the systems associated with their ideas, and, for each system, the factors that would influence the energy use of the corresponding construction activities. At least a triple of tags should be selected for each comment: one factor that influences at least one activity that is associated in turn to at least one system.

Since the test and Project A are based on the same case study, it is expected that both concept networks are similar. A correlation analysis between both networks can provide an internal validation for the proposed methodology to represent the network of the cases. We performed a QAP analysis and calculated the block analogy index $\bar{k}$ for the networks (the closer $\bar{k}$ is to 1, the more similar the networks are). The correlation factor found was 0.45 with a level of significance of 0.0002 (0.02%), and $\bar{k}$ was 0.51. Some of the factors that may have limited the results are the hypothetical nature of the test, the lack of experience of the students with O&G projects, and the time of the test.

Right after the test, we submitted an online questionnaire and conducted a focus group with the participants. In both evaluation methods, the participants assessed the concept map, the process of collaboration, the outcome of the test, and Green 2.0 as the interface. Overall, all the participants reported a high level of satisfaction with the four criteria. However, many commented on the difficulty of dealing with a limited number of tags. The issue of dealing with the tags may also have prevented from obtaining a better correlation/similarity between the networks.

4 Conclusion

Using unstructured data in a KM system has always been a problem in the construction domain. The industry struggles to deal even with intrinsic processes, such as documenting, historical productivity metrics, and benchmarking. In this sense, knowledge retrieval and representation methods such as lessons learned, case studies, and ontology have been widely used to improve the performance of projects. Nevertheless, these solutions have limitations regarding adaptability, applicability (capturing what is right for each project), documentation, and operation/maintenance (case studies and ontologies).

This research proposes a collaborative knowledge-based method to improve the use and management of unstructured data sources. It can be applied to the decision-making and planning processes related to energy management during the construction phase. Fundamentally, the proposed approach transfers text corpus into concept networks. A simplified taxonomy of factors, systems, and construction activities is used to capture the central concepts (to be used by the networks). The networks has edges (or links) set between the concepts (or nodes) that significantly affect energy use during the construction phase of a project.

The proposed method, as well as the system, can support the DM with better assumptions while providing a collaborative environment in which project teams can debate and co-investigate the best means to improve energy consumption during the construction phase.

Capturing knowledge from interviews, surveys and case analysis (the primary tools used in this work) is dependent on human interpretation, which may be subject to errors during the phases of profiling the answers and interpreting/collecting the relations. Despite its intrinsic disadvantages, the human interpretation allows for the exploration and detection of subtle relations in the participants' answers that are often difficult to perceive through other methods, such as surveys or questionnaires

[15,16]. The methodology to detect the relations cited above leads to semantically rich and contextualized knowledge, and the human interpretation is very effective in producing such outcomes.

Despite following a scientific procedure, detecting the relations from the case studies manually may be prone to errors. Future works can attempt to integrate automated knowledge retrieval tools to collect and represent the concept networks directly from the unstructured data generated in project environments.

Generalized blockmodeling has a significant number of parameters and constraints that influence the number of solutions. Although the number of inconsistencies in each solution is one of the most crucial parameters, each solution has to be manually interpreted before the one with the most substantive sense can be identified. Depending on the number of possibilities and solutions, this trial-and-error process may be overwhelmingly time-consuming. Therefore, the method to obtain blocks should be meticulously adapted to the needs of the study. For example, instead of using the four suggested clusters of activities, one can obtain a blockmodel from the three subcategories of factors of the concept map.

The analysis focused on the different types of blocks and the inconsistencies of each block to obtain valuable information regarding the case studies. Nevertheless, an alternative (and perhaps complementary) interpretation may be given to the opposite approach: What are the inconsistencies that have prevented blocks of the same type from being ideally equal? In addition, future works can attempt to implement semi-automated knowledge retrieval and representation tools to generate the cases and the networks.

References

[1] S.H. Liao, Expert system methodologies and applications-a decade review from 1995 to 2004, Expert Syst. Appl. 28 (2005) 93–103. doi:10.1016/j.eswa.2004.08.003.

[2] J. Read, Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification, in: ACL Student Res. Work., Association for Computational Linguistics, 2005: pp. 43–48. doi:10.3115/1628960.1628969.

[3] M. Park, K.W. Lee, H.S. Lee, P. Jiayi, J. Yu, Ontology-based construction knowledge retrieval system, KSCE J. Civ. Eng. 17 (2013) 1654–1663. doi:10.1007/s12205-013-1155-6.

[4] A. Singhal, Introducing the Knowledge Graph: things, not strings, Google. (2012). https://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html (accessed October 2, 2017).

[5] M.P.I.I. MPII, YAGO: Overview, Max Planck Inst. Informatics. (2014). https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/#c10444 (accessed October 2, 2017).

[6] R.R. Aragao, Using Network Theory to Manage Knowledge from Unstructured Data in Construction Projects: Application to a Collaborative Analysis of the Energy Consumption in the Construction of Oil and Gas Facilities, University of Toronto, 2018. https://search.proquest.com/openview/6d484bda470e83a7a7c2134d831b50c0/1?pq-origsite=gscholar&cbl=18750&diss=y.

[7] M.M. Bilec, A Hybrid Life Cycle Assessment Model for Construction Processes, University of Pittsburgh, 2007.

[8] C.L. Saunders, A.E. Landis, L.P. Mecca, A.K. Jones, L.A. Schaefer, M.M. Bilec, Analyzing the practice of life cycle assessment: Focus on the building sector Saunders et al. Analyzing the practice of life cycle assessment, J. Ind. Ecol. 17 (2013) 777–788. doi:10.1111/jiec.12028.

[9] S. Junnila, A. Horvath, A.A. Guggemos, Life-Cycle Assessment of Office Buildings in Europe and the United States, J. Infrastruct. Syst. 12 (2006) 10–17. doi:10.1061/(ASCE)1076-0342(2006)12:1(10).

[10] R.R. Aragao, T.E. El-Diraby, Using network analytics to capture knowledge: Three cases in collaborative energy-oriented planning for oil and gas facilities, J. Clean. Prod. 209 (2019) 1429–1444. doi:10.1016/j.jclepro.2018.10.346.

[11] D. Krackhardt, Predicting with networks: nonparametric multiple regression analysis of dyadic data, Soc. Networks. 10 (1988) 359–381.

[12] P. Doreian, V. Batagelj, A. Ferligoj, Generalized Blockmodeling, 1st editio, Cambridge University Press, New York City, NY., 2005.

[13] R.R. Aragao, T.E. El-diraby, Using blockmodeling for capturing knowledge: The case of energy analysis in the construction phase of oil and gas facilities, Adv. Eng. Informatics. 39 (2019) 214–226. doi:10.1016/j.aei.2019.01.003.

[14] T. El-Diraby, T. Krijnen, M. Papagelis, BIM-based collaborative design and socio-technical analytics of green buildings, Autom. Constr. 82 (2017) 59–74. doi:10.1016/j.autcon.2017.06.004.

[15] E. Rice, I.W. Holloway, A. Barman-Adhikari, D. Fuentes, C.H. Brown, L. a. Palinkas, A Mixed Methods Approach to Network Data Collection, Field Methods. 26 (2014) 252–268. doi:10.1177/1525822X13518168.

[16] A. Marin, Are respondents more likely to list alters with certain characteristics? Implications for name generator data, Soc. Networks. 26 (2004) 289–307. doi:10.1016/j.socnet.2004.06.001.