



Network analytics and social BIM for managing project unstructured data



Rodrigo Aragao^a, Tamer E. El-Diraby^{b,*}

^a Petrobras, Brazil

^b Ctr. For Intelligent Buildings Digital Twin, Dept. Of Civil & Mineral Engineering, University of Toronto, Canada

ARTICLE INFO

Keywords:

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

ABSTRACT

Unstructured data, mainly text in project documents and final evaluation or summary reports, is a container of tacit knowledge. We present an exploratory case study to use semantic network analysis to help capture and formalize elements of this knowledge. Text from project documents and chats by project participants over an interactive BIM platform were collected and arranged in the form of concept networks. With that, we used the rich literature in network science to formally study the networks. We illustrate the proposed approach by analyzing five concept networks. The case study illustrates the benefits of the approach—mainly, to develop a project-specific map of key concepts. Using network measures, such as centrality, identified the project's key issues and how they relate. Clustering measures identified possible knowledge constructs (interrelated concepts). Measures for quantifying the overall structure of the network can also be used to contrast projects.

1. Introduction

Unstructured text, such as online discussions, e-mails, and progress reports, are an essential component of any information or knowledge management system. Project text corpus is essentially an externalization of some of the tacit knowledge of project stakeholders. By default, this knowledge is context-sensitive: the knowledge contained in the discussions of project stakeholders is directly related and aware of the project needs and conditions. Hence, it is not surprising that, in many cases, unstructured data are more informative than other sources. This observation is particularly the case in the construction industry, where the knowledge tends to be subjective, context-sensitive, and multi-dimensional [1,2].

This research's primary objective is to explore the use of network theory to help create a more objective and consistent process for analyzing unstructured data (mainly text) in construction project management. The main idea for this case study was to use different means to transfer project documents and chats into concept networks. Then, using established literature on network analysis measures to externalize elements of the tacit knowledge contained in these documents. For example, what are the issues discussed, which issues are central to the project, how are project issues interrelated, and which groups of concepts cohere together in stable clusters? While this approach is ad hoc and not as formal as standardized knowledge management approaches [3], it has three contributions to the domain of knowledge capture and

re-use. First, because the nodes are a linguistic representation of topics/concepts, this makes them easily interpreted by humans. This approach is a major advantage over traditional product data models (such as IFC), which are mostly oriented towards computer use. Second, compared to content analysis, which does not necessarily produce a coherent set of topics, concept networks create an interlinked map of concepts. A network triangulates concepts, allowing users opportunities to understand their meaning. Third, concept networks are essentially a quasi-ontological representation of concepts. Both arrange concepts in a formal structure. However, concept networks have a superior benefit over ontologies: they are context-savvy. As suggested by this case study, they are produced through analysis of a project-specific text. As such, they reflect project-specific conditions. While ontologies provide a way to formalize knowledge, their static nature has shortcomings—mainly their insensitivity to context and the need for constant update.

Within a network, clusters of concepts, where a set of nodes are more densely connected to each other, can be detected. By contrasting these clusters across projects, we can detect patterns of association between concepts, which can qualify as knowledge constructs. Observing the changes in the network structure over the life cycle of a project can help analyze/study the dynamics of project knowledge evolution [4]. Which topics have gained more attention at different stages of project planning and construction? How have the sub-structures of concept connections changed with or around key project decisions? The overall network and the concept clusters represent a (unique) profile of the project, which

* Corresponding author.

E-mail addresses: R.aragao@petrobras.com.br (R. Aragao), Tamer@ecf.utoronto.ca (T.E. El-Diraby).

can be called the semantic signature of a project. Therefore, it is possible to use whole network similarity measures to compare and contrast the profiles of different projects. This approach can help stakeholders of a new project retrieve documents from previous projects that have the most relevance to the issues they are discussing for benchmarking.

This paper starts with details about the scope, objectives, and the design of the case study, including a summary of the projects used for analysis. This part is followed by a review of related works in information retrieval and knowledge capture. A review of the relevant network analysis measures is then presented. The next section covers the analysis of networks and how can they be related to project context. Then, we present an application on the use of blockmodeling for extracting possible knowledge constructs. The evaluation of the proposed approach is then addressed, followed by a set of conclusions.

2. Research scope

The direct objective of this case study is to explore the usability of concept network analysis in managing unstructured data and capturing some of its tacit knowledge. The study has been scoped along three dimensions: focus on developing the steps/methodology of the analysis; exploring the use of a no-model approach; and target the energy management domain. The aim is to focus on analyzing the suitability and benefits of such an approach/methodology, not on calculating its ROI. At the macro level, the formal assessment of ROI for information management system is a complex, possibly a multi-year econometric exercise. The ROI has been long established (see, for example, [5]); and is outside the scope of this study. At the level of a specific system, such as the one suggested here, the ROI evaluation should not be limited to quantifiable measures, such as the cost or time they save. Such measures are hard to assess within a project, due to the typical subjective nature of knowledge in the domain; and harder to contrast across projects due to the context-sensitivity of each project. The ROI is to be assessed based on supporting process improvement, capturing and re-using best practices, and improving manpower productivity and learning. The true value of information systems should no longer be “counted in only cash, bricks, mortar, and hardware; the true value in organizations today is how they do business [5]”. The contribution of this work is in developing and evaluating a viable methodology for conducting such an analysis. Therefore, this work does not aim to quantify the ROI of such a proposed approach at this stage.

A primary contribution of this case study is developing such a methodology based on a no-model approach. For long, our knowledge/information management system has followed a standards-based approach. These systems relied on a common benchmark for knowledge modelling, such as classifications systems, product data models, and ontologies. Standardized modelling is an intuitive process. Models serve two general purposes [6]: representing reality and specifying an idealized, desired system. However, no model can be as complex as reality. All models are wrong. We should not think that excessive elaboration will yield the correct one.

On the contrary, following Occam's razor, we should seek a minimal description of natural phenomena [7]. In normative thinking, models were central to common sense in science and technological domains. Some early purest claimed that even a bad model is beneficial. However, model maintenance can be a vicious cycle of ever-increasing complexity: sense changes in reality, abstract findings/observations, fit to the model, re-conceptualizing the model to accommodate any mismatches. Process-wise, model-free systems are based on simpler heuristics and are faster to train in a changing environment. Software engineering is a clear example of the success of no-model thinking. Recent progress in the domain has been made not by developing elaborate common models (i.e. UML graphs). Instead, adopting agile no-model approaches, such as SCRUM, has proven to be very helpful. As we move into a data-rich and data-driven world, we should not impose our models on reality. Instead, we should learn from patterns in real-world systems.

The no-model approach is linked to the belief in evolutionary knowledge production and co-creation. This thinking is “a new innovation paradigm where new ideas and approaches from various internal and external sources are integrated into a platform to generate new organizational and shared values. The core of co-innovation includes engagement, co-creation, and compelling experience for value creation. Thus, the practices of co-innovative organizations are difficult to imitate by competition [8]“.

The proposed approach could be applied to any text corpus. However, to reduce the complexity of analysis, we focused on one sub-domain of construction project management: reducing energy consumption during the construction phase of Oil & Gas facilities. The reason for this is twofold. First, avoid being too generic. We wanted the documents to belong to a coherent and specific domain within the broader spectrum of construction knowledge. Analyzing documents with overly diverse topics may dilute the findings of network analysis. i.e. a broad spectrum of documents will result in too generic concept networks that are sparsely populated. Only the most basic concept relations can be visible at this level. This viewpoint would have limited our abilities to detect any value or shortcomings in the proposed approach. Second, this domain deserves more attention. Much research work has been directed to energy analysis, savings, and management in the AEC industry. Limited work has been performed to optimizing energy use during the construction phase. The majority of research work has focused on almost every other stage of a facility lifecycle—for example, embedded energy analysis, pre-construction energy analysis, transportation-related energy analysis, and operational energy analysis. One hopes that using this specific domain as a target for analysis will present a contribution to the practice of both information management and energy management in the construction phase of projects. Again, the contribution to this domain is not in developing a model for calculating energy savings or suggesting an optimal process for doing so. Instead, it is in using the no-model mentality to help advance knowledge capture and sharing in this domain. Facilitating the flow of information and sharing knowledge in these large facilities, where construction may last years and includes large, energy-intensive equipment, can produce significant energy savings.

Specifically, the context used to illustrate this paper's proposed approach is to support capturing knowledge generated by project stakeholders while debating and planning the energy use in the construction phase of Oil & Gas facilities. Life cycle analysis (LCA) has been widely used to assess the environmental impacts of projects quantitatively. However, before “crunching the numbers”, LCA relies on the development of options, which includes several steps such as reviewing assumptions and collating related knowledge to represent data, processes, and system boundaries best. This process is accomplished by a team of experts such as planners, contractors, designers, suppliers. Because of the diversity of topics and the complexity of the issues, this team iterates in option development and exchanges (back and forth) arguments, problems, and suggestions for solutions. These exchanges contain valuable tacit knowledge. The question is if using semantic networks to represent project reports (or stakeholder communication) can be a suitable means to manage the exchange of information and knowledge.

Outline of the proposed approach.

Fig. 1 provides the basic scenario for the proposed system (the numbers in the figure correspond to those below).

1. During a construction project, teams of engineers, contractors, and other stakeholders collaborate and negotiate options for the design or construction of a facility. This process can take place during the planning, design, and construction stages. First, they have to deliberate the “what”: are construction approaches, work strategies, and resources feasible? Second, the “how”: how do project conditions (technical, organizational, environmental) impact energy use, and how can it be reduced? These interactions represent a map of

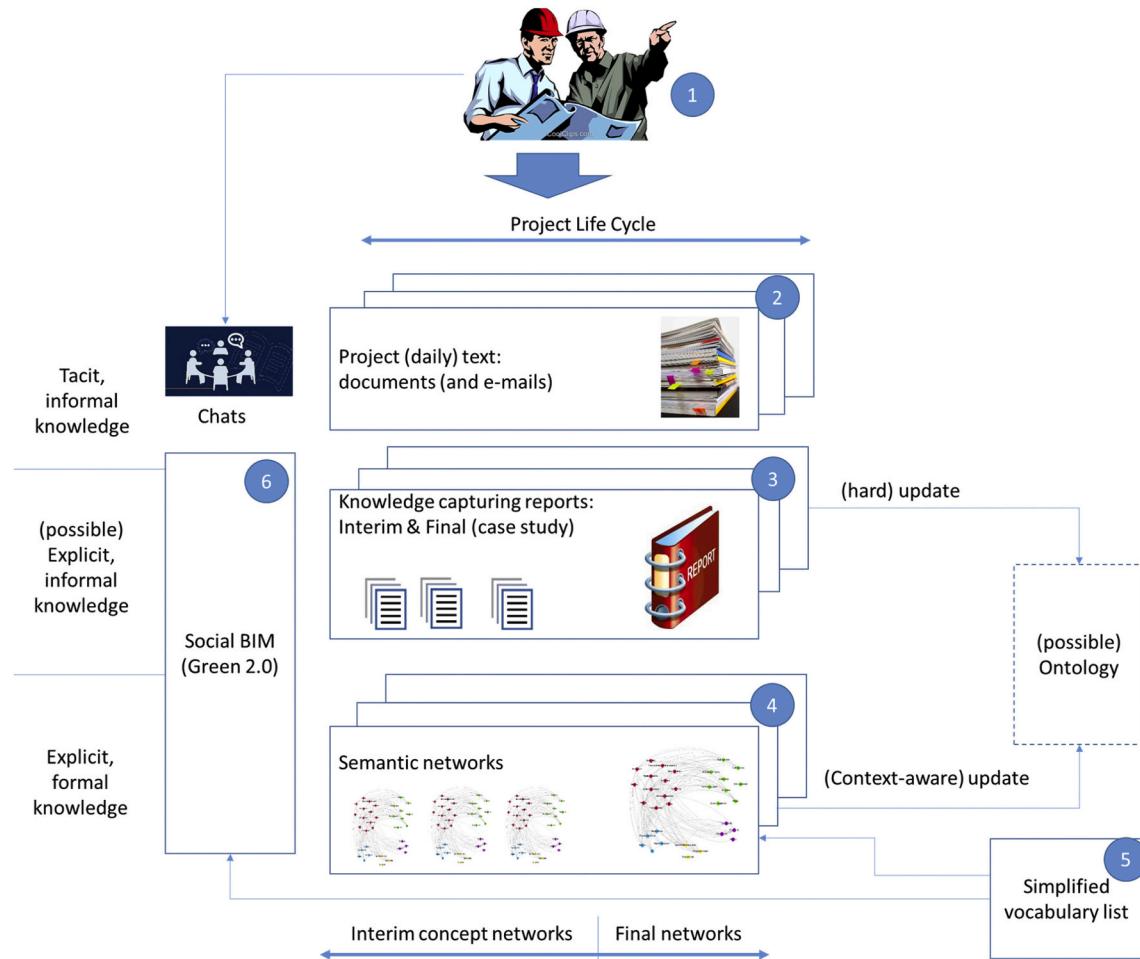


Fig. 1. Scope and Context.

positions, arguments, and decisions made by stakeholders. Their deliberations reflect their expertise and are expressed in the form of key knowledge concepts. However, this knowledge is tacit and informal.

2. The organization, interested in capturing this knowledge, tracks documents (text) containing/ summarizing their deliberations.
3. Frequently, a set of project reports are produced. These reports summarize critical project conditions, major decisions, and the justification for them. At the end of a project, a final report is produced as part of the company's knowledge management program. These reports, stored in a database, make the knowledge exchanged during the project deliberations/communication explicit. They discuss issues, ideas, problems, options, arguments, and lessons learned. Future projects are encouraged to review these reports to benchmark best practices.
4. The knowledge contained in these reports, while explicit, is not formalized—or at least computer-processual. Retrieving reports in the form of “free text” is inefficient. This challenge is where the proposed approach is valuable. The interim and final reports are transferred into concept networks. Using network analysis measures, stakeholders can study influential concepts and their interrelationships. An interim network is developed for a new project and can be matched more accurately to previous, similar benchmarking projects. The contents of the reports can be transferred into concept networks through a manual or semi-manual process. In the first method, a knowledge manager, working with project teams, can create the network based on the analysis and discussions of the main report/project concepts and their interrelationships. Alternatively, a

content analysis software can be used to generate topics from reports [3]. Some interactive BIM-based platforms can be used to generate networks automatically (see the section below).

5. A simplified vocabulary can be used. i.e. to avoid confusion and to create uniformity in the concepts of the semantic networks. The networks are expressed using a common lexicon. This lexicon can be as simple as a list of terms, a taxonomy, or a full-fledged ontology.
6. One of the evolving, relevant sources for knowledge is chats between stakeholders. It is not easy to include this source of knowledge in the traditional knowledge formalization approach (i.e. report development). In this research work, we utilized the social benefits of BIM to capture and formalize chat concept networks. Green 2.0 is a BIM-based platform that allows users to co-comment on any issue related to the project. The analytics engine behind Green 2.0 uses network theory to generate a semantic network of user's comments.

Finally, if the organization uses an ontology, it is very hard to use the knowledge gained from final reports to update it. With the semantic network, several network analysis measures exist to capture patterns of concept association across projects. These clusters of concepts represent possible knowledge constructs that can be added to the ontology.

It is fundamental to point out two scope issues. First, the proposed approach is semi-manual. It is expected that a team will work on securing the documents and transfer their text into a proposed or initial network to seed the analysis by project stakeholders. The final network will evolve and be approved by them before the network analysis. This procedure is a reflection of the most fundamental objective of this

approach: help people streamline their analysis of issues, their debates, and their information exchanges. Consistent with no-model thinking, the objective is to support knowledge exchange and co-learning amongst project stakeholders, not to tell them what they should “know.” Second, we did not aim to exhaustively examine the use/study of all possible network analysis measures, let alone finding the measures that yield higher performance or optimal results. There is a significant number of these measures (centrality alone has close to 100 types); and we only have three real-world cases.

More importantly, the analysis is context-sensitive. This is not just a reflection of the project boundary conditions, but also the composition and expertise of the stakeholders who are creating and analyzing the relevant knowledge. Relying on a self-regulating team in setting up the analysis parameters, selecting the measures, and interpreting the results is more effective and sustainable. This reliance is due to the increased importance of boundary conditions in shaping information management systems and defining what is valuable knowledge. “Self-regulating groups are more task effective to the extent that members have: (a) a moderately differentiated task; (b) high boundary control; and (c) high task control [9].”

In summary, the proposed approach falls into the domain of information management and retrieval. However, at the same time, it aims to do so with a focus on capturing and using knowledge constructs. In other words, while document profiling and retrieval are the domain of information systems management, doing so in the form of semantic networks creates an environment for capturing and reusing knowledge. This scheme achieves three goals. First, horizontal integration: generating

network from project documents (interim and final) captures all stakeholders’ views (across their diversified disciplines). Second, vertical integration: repeated use of this approach over a specific project’s life cycle can help track the evolution of issues as the project progresses. Third, longitudinal integration: repeated generation and analysis of project concept networks across projects help capture knowledge constructs from information exchanges.

3. Research design & steps

The research work included six main steps as shown in Fig. 2. They are summarized below. Subsequent sub-sections provide further details about three of these steps:

1. Develop a simplified common vocabulary (lexicon). Through reviews of relevant literature, a base concept map was developed. It is a simplified taxonomy of concepts associated with the energy use during the construction stage of Oil & Gas facilities. Reusing the terms of this lexicon ensures the networks are always represented in the same language. A survey of 112 experts was conducted to a) assess the relevance of the suggested concepts, and b) to establish relationships between them. In total, 34 concepts were included in the lexicon (see next sub-section for more information). It is worth noting that the lexicon was intentionally kept small to illustrate the power of network analysis. In other words, if network analysis can produce meaningful results with a limited set of terms, this showcases the value of portraying concepts in the form of a network. In

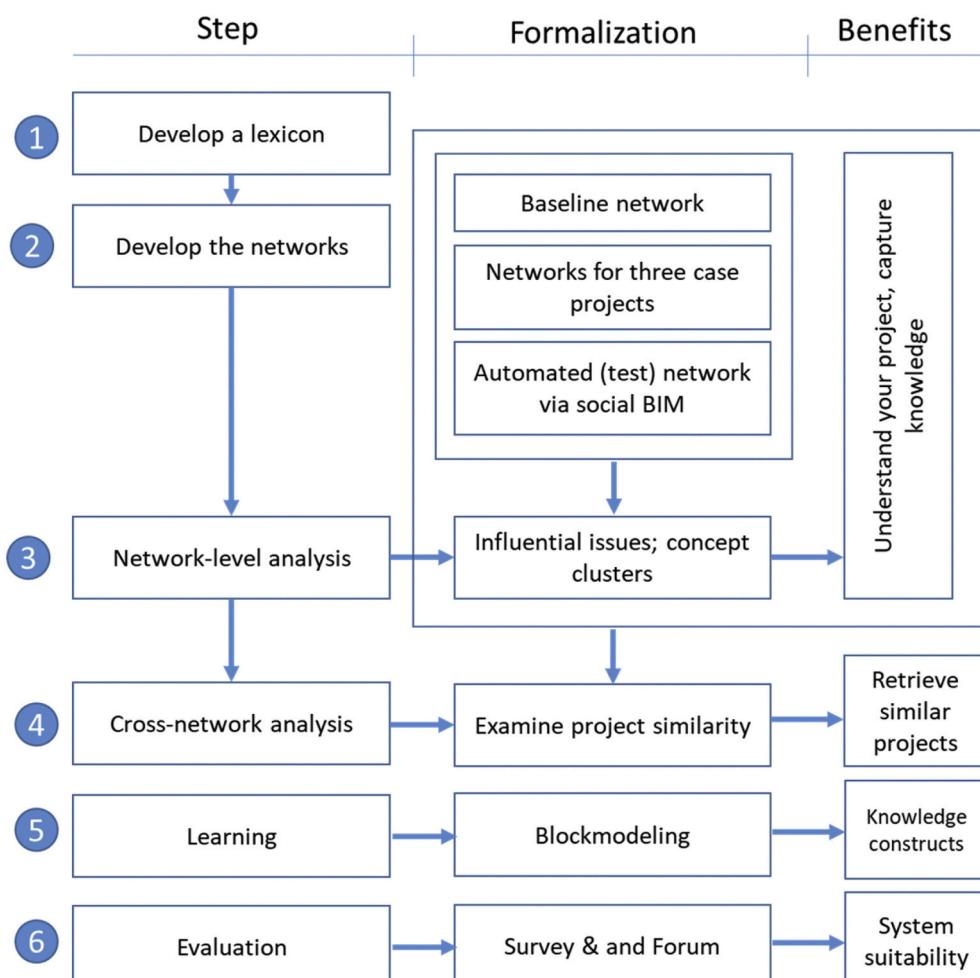


Fig. 2. Research design and steps.

- addition, the lexicon is not claimed as a formal representation of knowledge in the domain—just a simple list of terms to facilitate uniform comparisons between the networks.
2. Developing the networks. We developed five networks in this case study:
 - Baseline network (BL network): a 34×34 adjacency matrix was established for the concepts of the lexicon. Out of 1156, possible relationships. Some relationships were illogical, so only 573 relations were used. Experts were asked to evaluate the viability of these relationships. If more than two experts agreed that a relationship exists between two concepts, a link in the network was established between these two concepts.
 - Sample projects: we attained text about three actual projects. They all are oil and gas projects in Brazil (hereinafter referred to as projects A, B, and C). Shortly after they were finished, the first author interviewed 7–8 stakeholders from each project. The stakeholders' input was combined into a “final” report for each project (in each case, stakeholders had the chance to amend drafts of the reports). In the interviews, the participants responded to open-ended questions about the challenges, best practices, opportunities, and risks that affected the energy use during the construction phase of their project. A relation or link was established whenever the participants communicated a semantic association between two concepts. These relations are transferred to an adjacency matrix, and a network of concepts was generated for each case project.
 - Green 2.0 project (test/G network): a fourth project was created through the simulated reincarnation of Project A. Graduate students with field experience, as well as participants in Project A, were invited to act as if they were stakeholders for (a new) Project A. They used Green 2.0 to discuss the design and construction plans. They debated construction plans, constructability issues, safety, cost, schedule, energy savings, and management. The platform developed a semantic network for their deliberations.
 3. Intra-network analysis: for the five networks, we utilized a set of network measures to help explore their unique features: which concepts were influential and how clustered the networks were.
 4. Inter-network analysis: network similarity measures were used to compare the five networks.
5. Learning through capturing patterns: blockmodeling was used to study the networks to help extract knowledge constructs (cohesive subsets of each network).
6. Evaluation: the proposed approach was designed to help illustrate and study two fundamental dimensions of unstructured data management: capturing knowledge from unstructured data; and formalized analysis of such knowledge. Participants in the simulated project were surveyed, and a forum was conducted with a group of them to assess how useful was the proposed approach in achieving these two goals.

3.1. The lexicon & the baseline network

Concepts related to the construction of O&G projects were presented along three dimensions, as shown in Fig. 3. These concepts can be linked to IFC. However, the lack of IFC models for O&G did not allow this for our case.

1. Construction activities: 13 of the main tasks for construction work in an O&G project. These range from excavation to welding to scaffolding.
2. Systems: these are physical sub-components of the project. To an extent, they represent the specialties in the project work (nine in total). For example, civil works were distinguished from electrical, and both were distinguished from instrumentation.
3. Factors: these are listings of the boundary conditions of the project. They are categorized into three main groups (12 in total):
 - Design factors: this refers to the assumptions, prevailing codes, and objectives of the project as seen/developed in its pre-construction stages.
 - Site characteristics: this encompasses site attributes and constraints, such as weather, soil conditions, and location.
 - Resources: this refers to the type, capacity, and productivity of the resources used in the construction stage.

An adjacency matrix ($34 \times 34 = 1156$) was used to transfer the lexicon into a semantic network. Not all 1156 possible relationships were logical—for example, scaffolding and geographical location; underground pipelines and masonry. The number of suggested logical relationships was 510. However, surveyed experts had the opportunity to override the

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

Fig. 3. The proposed concept lexicon.

initial assessment by the authors. They added 63 more, for a final total of 573.

3.2. The three projects

In addition to the baseline network, we investigated four case projects. The first three (projects A, B, and C) are real-world projects. The fourth is a hypothetical repeat of project A. The following is a summary of the projects.

Project A: the project is a Diesel exhaust fluid (DEF) plant. DEF consists of an aqueous solution of 32.5% of urea and deionized water, and is consumed during the selective catalytic reduction (SCR) process, whereby it is injected into the exhaust systems of diesel engines. The project is an add-on to a large existing fertilizer factory. The project was completed on schedule. However, the measures taken during the construction phase to reduce energy usage were limited.

Eight people who participated in the project were interviewed. Participants cited the following lessons learned. The use of a temporary shelter positively affected the energy use in most activities because it reduced the impacts of the rainy season, especially during the tank assembly. However, the mobilization of extra resources to make up for delays in the schedule was very counter-productive and energy-intensive (acceleration plans).

Project B is an offshore liquefied natural gas (LNG) terminal constructed to import and re-gasify LNG in Brazil. The project included building a jetty with the main platform, ten mooring dolphins, and the respective linking footbridges; two automated loading arms; a metering station; an electrical substation building; a hydrogen and compressed air station; two firefighting cannon towers and the pump area; a instrumentation/control system; and the launching of an 8-km electric-optical cable, responsible for the power supply and communication of the terminal with onshore installations. Seven project participants were interviewed.

The initial contract term was one year and ten months. Initially, the term was considered quite feasible. Unfortunately, however, the weather conditions and rough sea during the winter hampered the project completion substantially. Weather, sea conditions, and a longer than expected time to obtain the Brazilian Navy's clearance caused a delay of four months in the foundation of the terminal. Nevertheless, the project was expedited and concluded on time. Some relevant facts that influenced the energy consumption included the following. The sea conditions impacted the logistics and deep foundation works; large scale prefabrication of concrete parts and pipe spools was an effective strategy used; unfavourable sea conditions and weather delayed rigging (material handling); the flawed seabed investigation during the basic design phase impacted the deep foundation services, causing delays in the project and therefore increasing the energy demand. Finally, as suggested by the contractor, a last-minute change in the design, which reduced the number of piles and increased the diameter, saved energy and time because it reduced the number of moves for the boats and barges during the drilling and concrete preparation. However, to guarantee the project schedule, the contractor had to mobilize extra resources. Most interviewees agreed that the energy spent on transport could have been optimized if the logistics planning had also integrated energy solutions instead of just cost and schedule.

Project C was an onshore section of the 45-km gas pipeline responsible for transporting the regasified gas from the LNG terminal, discussed in Project B. The project included a 30-km onshore gas pipeline; a 15-km offshore section; an intermediate pressure reduction station; and a flow control station to connect the gas pipeline with the national NG network at its end. The delay in the terminal clearance deferred the pipeline start-up despite its completion on time. Yet, the construction went through several problems that put the whole project's schedule in jeopardy: financial problems incurred by the contractor and delays to finalize the hot tapping. Hot tapping is the operation to connect a new pipeline with the existing one in operation. This activity is a critical

operation since the pipeline in operation is submitted to welding and boring services.

Seven project participants were interviewed. The contractor postponed the construction activities due to the rainy season, which may have caused reworks, stoppages, low productivity, and increased energy spending, especially for earthworks and excavation. To expedite work, skidded equipment was delivered as modules to the site. Hot tapping was a challenge. There was a lack of information about existing pipelines and other buried structures in the right-of-way or the aboveground stations (brownfields), leading to low productivity of the resources, delays, and an increase in energy use. Pipeline construction is heavily dependent on the implemented logistics. Therefore, besides the equipment's direct energy consumption, other factors that are difficult to predict indirectly influence the outcome, including the site location, type of soil, the productivity of the resources, and, most importantly, time to completion. Capturing these factors can be very helpful for future designs.

A hypothetical repeat of project A was created on the BIM-based platform called Green 2.0 [10]. Green 2.0 is equipped with a social-media-like interface. It allows users to visualize 3D models of projects, where they can select a project component (IFC product) and comment on it. Green 2.0 captures comments and responses to them (and stakeholders who provided them). The platform then uses a series of network analyses to portray social ties, semantic networks, and sentiments on the participants' topics.

We invited four experts who participated in Project A and six civil engineering students to act as the project team, deliberating on the energy use during the construction phase. The participants had a chance to read several project documents and the report generated for Project A. Students also attended a kick-off meeting that provided basic background about the project context. Each participant was given a role in this "test case." The roles were chosen to guarantee the involvement of the three main project stakeholders: owner, owner representative (consultant), and contractor. Once the participants were comfortable with the project's information, they were asked to debate the project on Green 2.0. This simulated project environment lasted for 34 days. The objective of this test case was twofold: (i) compare the network generated for project A by the participants and that generated manually by the authors, and (ii) evaluate the collaboration process, the suitability, and ease of use of the proposed approach.

In Green 2.0, the user starts with selecting a system or an IFC product. After commenting, the user can tag it. There are three categories of tags to choose from: systems, activities, and factors (from the lexicon shown in Fig. 3). For example, let us assume that "piping" and "civil structures" are the two systems tagged by the user. For the "piping" tag, the user may select specific activities such as "welding" and "piping assembly"; and for "civil structures," the user can select other relevant activities such as "deep foundation" and "concrete preparation." The user must also select at least one impacting factor associated with/influencing each selected activity.

4. Related work

Early on, in the 1990s, a set of challenges of managing project information and unstructured data were identified [11]. They remain valid today. These include the need for a system for information exchange that achieves three main tasks: 1) horizontal integration across the diversified disciplines of stakeholders; 2) vertical integration to match the evolution of issues along the stages of a project life cycle; and 3) longitudinal integration over time. The latter is the bridge between managing information flows and capturing patterns in those flows that can serve as "knowledge constructs."

In practice, knowledge is contained and reused in different formats. One of the key forms are lessons learned. These are typically presented as enumerated rules or lists of best practices. They are also increasingly presented in the form of free-text documents. However, the subjective

and multi-disciplinary nature of construction knowledge entangles the extraction and agreement on reusable rules [12]. More importantly, the context-sensitive nature of projects makes it difficult to define which of the written rules match project scope or conditions [13]. While documents are more superior in capturing meaning and context, the need for manual/human reading and inference makes capturing knowledge in the form of free text a complicated task concerning knowledge management—such as capturing, formalization, and retrieval. Searching for documents that have the most relevance to the needs of a new project is challenging. In short, the complexity of construction knowledge makes unstructured data (such as text) essential to adequate capture of lessons learned and knowledge. At the same time, this free-text format makes it hard to manage such knowledge [14].

With the increased volumes of information, the growing complexity of their flows, and the increasing reliance on them, an extensive number of attempts (research work and software systems) aimed to automate the processing of documents. Initial attempts relied on benchmarking a document against a standard static vocabulary, with varying levels of sophistication: classification systems, data models, taxonomies, and ontologies. These structured formats suffered from the problems of model-based systems, which reduced their efficacy [15–17]. Modes of model-free text mining were used to overcome some of these restrictions. For example, Caldas et al. [18] implemented the earliest attempts to use text mining to retrieve construction documents. Based on the content of a document, a set of classes is assigned to it. Through matching classes of two documents, their similarity can be assessed. A supervised machine learning algorithm is used to “learn” or identify the classes based on the frequency of their appearance in a set of target documents. A significant drawback of free-form text mining is that the classes used are unrelated—they are treated as a bag of concepts. A seed classification system can enhance the learning process and create a base for consistency in the classes used.

Nevertheless, the need is to search for documents with relevant knowledge, not just similar keywords. Counting and comparing bags of keywords neglect the complex interrelationships of concepts. One of the principal solutions to this problem is to use classes arranged in an ontology [19,20]. However, the ontology’s static nature (and almost all other data models) is a considerable shortcoming of all systems that rely on them [21,22].

Semantic networks can be used to map entities/concepts from a text along with their interrelationships. Network analysis measures can then be applied to study the maps of concepts. Degree centrality, for example, counts the ties of a node. A high centrality may refer to fundamental nodes. However, a node located in the paths that connect too many other nodes (betweenness centrality) could be more critical. A node that has a lower number of connections could still be influential if these are connections to important nodes, referred to as Eigenvector centrality [23]. Instead of node-level analysis, clustering aims to detect the relationships between groups of nodes. Clustering measures offer additional insights. Clusters are nodes with a higher level of connections amongst themselves than to the overall network. These tightly connected nodes can be seen as a knowledge construct or a mini ontology. Another set of measures can be used to qualify and study the whole network.

For example, the shortest path length is the number of links that separate any two nodes in a network. The average path length (L) is the mean of all shortest path lengths. Lower values of L indicate a dense network (overly connected network). The average clustering coefficient is the mean for the clustering coefficient for each node and reflects the modularity of the network [24]. More advanced measures can be used for cross-network. Two nodes are structurally equivalent if they are connected to the same (exact) nodes. Regular equivalence, a relaxation of the structural equivalence, refers to nodes that share links to nodes that are themselves equivalent [25]. Expanding this to the network level, we can measure the degree of similarity of two networks. However, similarity measures are node agnostic. i.e. they do not recognize the specifics of which nodes are connected to each other and its

implications. Blockmodeling is a more sophisticated approach for evaluating network similarity. It is more cognizant of which specific nodes tend to be more associated together. In blockmodeling, the nodes of each network are arranged in an adjacency matrix [26]. A block is defined when a cluster of nodes has the same pattern of ties in both networks. They are found through a process called permutation [27]. The number of blocks (matched clusters in both networks) highly depends on the cluster’s size, which is iteratively set by the analyst. Hierarchical clustering aims at grouping nodes based on the Euclidean distance. Other algorithms, such as STRUCTURE and CONCOR, use node equivalency [28]. Blocks essentially are small coherent groups of concepts that tend to appear and associate with each other. Effectively, they can be interpreted as mini ontologies or potential formalized knowledge constructs.

In generalized blockmodeling, the user pre-selects block types and then specifies that a block is found if there is a complete one-to-one match between *all* the nodes in both networks’ clusters. This rule can be relaxed by accepting that the number of connections in *each row* (row-regular) or *each column* (column-regular) is the criterion to identify a cluster of nodes to form a block. Finally, the user may be just satisfied if the number of *any* node connections in both blocks are the same, called a regular block [25]. The purposeful nature of generalized blockmodeling provides the researcher with direct influence, deciding the degree of similarity and how nodes shall be connected [29]. The main shortcoming of blockmodeling is that it is computationally complex, can take a long time and demands frequent users’ interferences and analyses to make sure the final result is meaningful [30].

Transferring the deliberations of experts into a network can be done through different means. The most basic and, possibly, the most accurate, especially with shorter documents, is a human interpretation. A specialist can review documents or interview authors of a document. Despite its slow-pace nature, manual generation of networks allows for exploration and detection of subtle relations between participants’ answers that are often difficult to perceive in other methods [31]. To handle the subjectivity and dynamics of knowledge in the field and to accommodate users, a system that can capture/track their exchanges and use text analytics to formalize the knowledge in their interactions can be a win-win. It will be easy for users, uniquely fitting to the subjectivity and context-sensitivity of construction and will support the indirect formalization of knowledge. Promising new automated tools are available—for example, Google’s Knowledge Graph [32], OPEN IE [33], and KnowItAll [34]. Natural language generation is an emerging technology to automatically transform texts into simple reports with the synthesis of key takeaways [35].

However, the construction industry is notorious for the low interest in pro forma tasks such as documenting or even developing as-built drawings, let alone taking the time for developing case studies or synthesizing lessons learned [4]. The difficulty of gaining interest is compounded if users need to customize the analysis, software, or the ontology [36]. Specific to the construction industry, social BIM can be a suitable solution to the problems. Social BIM is a socio-technical mode of BIM that supports the co-creation process. At one end, co-creation refers to the collaborative, real-time development and management of BIM files, based on the visual and interoperable capabilities of BIM. It enables ‘shared situational awareness’ by empowering remote participants with visual and remote control of BIM models using ‘groupware’ [37]. On another end, in transformative planning, co-creation refers to stakeholders’ empowerment to innovate and generate new knowledge. It is based on the assumption that knowledge is distributed and evolutionary. All participants have a contribution to the development and the very definition of what knowledge is, and they iteratively collaborate in the process of co-learning and validation of knowledge [38].

The push for considering social aspects, collaborative processes, and evolutionary systems in developing and managing knowledge emanates from juxtaposing the typical work, technical processes and tools on one side; and the complexity of personal, behavioural and social forces on another side [39]. Leading initiatives in this regard, the Tavistock

Institute promoted actual research in the workplace [40]. This approach included relying on industrial psychology, field theory, and open system theory to develop knowledge management systems based on emergent synergies between person and practice [41]. Some social BIM platforms integrate social-media features with BIM. They enable commenting directly into the BIM file. For example, Green 2.0 [10] establishes a user interface on top of BIM to navigate the design. It also allows users to comment on each product. An automated social and semantic network is created periodically from their interactions, with limited user overhead.

One of the areas that can benefit from better management of unstructured data is the domain of energy use during the construction phase. The analysis is multidisciplinary, context-sensitive, and have not received equal attention from researchers. During the construction phase, energy consumption is typically considered a percentage of the construction materials' total impact. Estimates range markedly: from approximately 6% to 30% of the manufacturing energy [42]. A comprehensive framework using a hybrid LCA approach of an office building and a precast parking garage [43] showed that the energy values were significantly larger than reported by other researchers—for example, the findings were ten times larger than those reported earlier [44]. These uncertain figures highlight the importance of deliberation ahead of crunching the numbers. A literature review of LCAs in the building industry emphasized the need for using stochastic studies because of the strong influence of project context and the assumptions made on the results [45].

5. Network-level analysis

Five networks were available for analysis. A multitude of network analysis measures can be used to deduct key features of the text and the knowledge behind it [46]. Table 1, Table 2, and Table 3 show four centrality measures for three networks: the baseline network, case projects B and C (significant values are bolded in the Tables). Analyzing centrality is fundamental for capturing essential project issues. For example, in Project B, betweenness centrality shows that transportation, rigging, and masonry work were the most central activities. They are in the paths connecting most other activities. In other words, most activities are connected (sequenced) through paths that go through these three activities. They are the most significant, always on a critical path. Out-degree centrality shows that the geographical and sea attributes were the most important factors in the project—these factors are connected to the largest number of nodes. This observation was expected as the project is an offshore facility. The values of in-degree centrality have high variability. However, most of the construction activities have high in-degree. This broad range of values may indicate a complex project design and or plans: each activity receives input, constraint, or influence and is connected to too many other issues. Eigenvector centrality points out that there is a higher importance to piping, utilities, and site characteristics. They are connected to the top influential nodes.

As expected, the logistics of offshore projects is the most influential factor (Project B is an offshore terminal). However, masonry work, which is typically not associated with high impacts on energy consumption in O&G projects, has the third highest betweenness. It is possible that the large scope of civil structures (a type of system), which has the second highest in-eigenvector centrality, may have contributed significantly. These structures, such as the jetty (a structure that resembles a pier in the sea), are large and complex. In this project, they included a considerable demand for masonry. Masonry, in this case, is an important activity not because it is an energy-intensive activity but because it affected the performance of the construction of a critical system (for example, the jetty).

Another example of the use of centrality in uncovering project issues was captured in the Project C network. The “utility sets” system had the third most prominent in-eigenvector centrality. The idea of utility sets being a critical system in a pipeline project is counterintuitive. This system is typically not as complex in these types of projects. Yet, Project

Table 1
Centrality for the baseline network.

Nodes	Out-degree	In-degree	Betweenness	In-Eigenvector
Construction Activities				
Onsite transportation	12	17	68.392	0.784
Piping assembly	13	16	48.753	0.857
Rigging	10	13	35.613	0.723
Pipeline assembly	11	18	33.901	0.91
Welding	13	8	26.067	0.544
Deep foundation	10	12	18.437	0.445
Concrete preparation	9	12	16.134	0.503
Earthworks	9	11	13.825	0.561
Hydrotesting	5	7	12.878	0.305
Excavation	10	13	11.668	0.562
Scaffolding	1	3	7.059	0.136
Electrical/instrumentation assembly	2	6	6.863	0.441
Masonry	2	6	0	0.273
Physical Systems				
Civil structures	3	9	11.736	0.615
Underground pipeline	3	7	8.62	0.437
Piping	3	7	8.62	0.4
Utility sets	2	6	5.613	0.388
Dynamic equipment	3	6	24.876	0.382
Building/shelter	3	5	7.058	0.315
Static equipment	3	2	1.782	0.132
Electrical equipment	1	0	0	0
Instruments	0	0	0	0
Design				
Design parameters	17	14	90.091	0.348
Type of material/equipment	16	15	105.541	0.348
Standard/Code	7	13	19.927	0.311
Site Characteristics				
Sea conditions	17	0	0	0
Weather	16	0	0	0
Geographical location	16	0	0	0
Soil	15	0	0	0
Adjacent structures	15	0	0	0
Resources				
Type/Capacity	15	20	44.596	1
Productivity	15	18	35.21	0.918
Consumption	12	17	18.33	0.831
Quantity	11	19	26.412	0.928

C was different. It was a project to build a relatively short 30-km pipeline, which included two complex stations and the installation of several large pieces of equipment to process natural gas. The complexity of the “utility sets” system, within a relatively short pipeline project, pushed it to be relatively more important. In this case, network analysis was sensitive to the project “context.”

For Project B, onsite transportation had the highest in-eigenvector centrality. i.e. it is connected to highly influential nodes. The team reported in the project records that this was actually a significant issue. Then, they decided to buy and haul much of the equipment in large semi-assembled skids. This measure drastically reduced the scope of offshore, onsite construction activities.

6. Cross-network analysis

Can we compare different projects? Which two projects have the most similar conditions/features? A simple approach is to compare the structures of the two networks. Centralization is a measure to characterize the overall structure of a network. It assesses how central are the high centrality nodes. i.e. it compares the sub-networks of central nodes in two networks. Table 4 below shows the degree centralizations for each of the three project networks. In general, out-degree centralization still hovers around 0.5. Not much can be deducted from this. However,

Table 2
Centrality measures for Project B.

Nodes	Out-degree	In-degree	Betweenness	In-Eigenvector
Construction activities				
Onsite transportation	19	9	54.889	0.103
Rigging	8	10	23.973	0.167
Masonry	7	9	19.156	0.167
Electrical/instrumentation assembly	4	9	13.050	0.373
Welding	4	9	11.239	0.373
Concrete preparation	6	9	10.373	0.167
Piping assembly	2	9	5.317	0.373
Hydrotesting	3	4	1.850	0.103
Deep foundation	1	9	1.589	0.167
Excavation	1	8	1.339	0.167
Scaffolding	5	2	0.892	0.127
Earthworks	0	0	0.000	0.000
Pipeline assembly	0	1	0.000	0.064
Physical systems				
Utility sets	0	7	0.000	1.000
Civil structures	0	8	0.000	0.888
Piping	0	6	0.000	0.770
Electrical equipment	0	5	0.000	0.579
Instruments	0	4	0.000	0.500
Static equipment	0	4	0.000	0.461
Building/shelter	0	5	0.000	0.451
Dynamic equipment	0	3	0.000	0.397
Underground Pipeline	0	0	0.000	0.000
Design				
Design parameters	10	4	24.500	0.000
Type of material/equipment	1	0	0.000	0.000
Standard/Code	0	0	0.000	0.000
Site characteristics				
Geographical location	14	0	0.000	0.000
Sea conditions	13	0	0.000	0.000
Adjacent structures	4	0	0.000	0.000
Soil	3	0	0.000	0.000
Weather	3	0	0.000	0.000
Resources				
Productivity	11	6	20.417	0.103
Quantity	11	2	0.500	0.000
Consumption	10	2	2.917	0.064
Type/Capacity	6	2	1.000	0.000

in-degree centralization, while consistent, is much lower. This measure indicates that a few nodes with way too many out-relations linking them to too many nodes. The exception is Project A. All three centralization measures are relatively close in value. This can refer to a uniformly connected network. The chances of having extremely influential nodes in this network are possibly low.

Alternative network-level measures can be helpful along with centrality in project-level comparisons. Table 5 shows the average path length (L) and the average clustering coefficient (C) for the baseline network (BL). These values are also compared to the mean values of an equivalent random network. Benchmarking the values of L and C against those of a random network can be informative in qualifying a network, which is a common approach in network analysis. The results show that the baseline network has a comparable length (L) to the random network. However, there is a significant difference in the value of the clustering coefficient between the two networks. This remark means that the baseline network has smaller sub-networks (clusters of nodes) that are coherent enough to stand-alone—making them worthy of further analysis. The values for L for the three project networks were smaller than the corresponding random networks. Small-world networks typically have L higher than their equivalent random networks [17]. This result suggests that the case networks may not be “small-world” networks, which is a feature that indicates network maturity—it matches

Table 3
Centrality measures for Project C.

Activities/factors	Out-degree	In-degree	Betweenness	In-Eigenvector
Construction activities				
Onsite transportation	17	9	48.746	0.181
Welding	5	10	13.846	0.279
Rigging	7	8	10.796	0.279
Pipeline assembly	2	10	7.063	0.332
Concrete preparation	4	8	6.846	0.181
Excavation	5	8	6.496	0.181
Earthworks	8	8	5.796	0.181
Electrical/instrumentation assembly	3	6	4.500	0.181
Masonry	3	6	2.625	0.181
Piping assembly	2	8	2.033	0.279
Hydrotesting	2	2	0.250	0.098
Deep foundation	0	0	0.000	0.000
Scaffolding	0	0	0.000	0.000
Physical systems				
Piping	0	9	0.000	1.000
Underground Pipeline	0	8	0.000	0.931
Utility sets	0	6	0.000	0.750
Civil structures	0	6	0.000	0.643
Building/shelter	0	6	0.000	0.643
Instruments	0	4	0.000	0.529
Electrical equipment	0	3	0.000	0.348
Dynamic equipment	0	0	0.000	0.000
Static equipment	0	0	0.000	0.000
Design				
Type of material/equipment	10	0	0.000	0.000
Design parameters	10	3	2.667	0.000
Standard/Code	0	0	0.000	0.000
Site characteristics				
Soil	11	0	0.000	0.000
Geographical location	11	0	0.000	0.000
Weather	10	0	0.000	0.000
Adjacent structures	6	0	0.000	0.000
Sea conditions	0	0	0.000	0.000
Resources				
Productivity	10	3	21.500	0.152
Type/Capacity	5	1	0.000	1.833
Consumption	1	1	0.000	0.000
Quantity	1	0	0.000	0.000

Table 4
Comparing the centralization of the three case project networks.

Project Reference	Degree Centralization	Out-degree Centralization	In-degree Centralization
Project A	0.343	0.332	0.208
Project B	0.473	0.459	0.178
Project C	0.421	0.409	0.190

Table 5

The average path length (L) and average clustering coefficient (C) of the baseline network.

Measure	Baseline network	Random (mean)	CI-95% (random networks)
L	1.819 (+1%)	1.797	[1.79, 1.80]
C	0.407 (+52%)	0.268	[0.26, 0.27]

what is expected in a thriving “small” world. With all four networks built using the same lexicon, this indicates that the baseline network is more coherent (and well-built) than any of the projects’ networks. More work and analysis is needed to increase the depth of the projects’ networks. Checking L and small-world behaviour, in general, can serve as a quality check for a knowledge manager in a company. The manager can reject

networks that do not portray the small-world phenomenon or request that they are redeveloped.

More advanced measures have been used to contrast networks more specifically. These are typically referred to as structural measures. For example, the network dissimilarity measure posits that two nodes are similar (or equivalent) if each is connected to the same nodes. In this case, they are structurally equivalent (or 100% similar). If each node is connected to entirely different sets of nodes (there is no overlap in the nodes they are connected to), they are assigned a similarity measure of 0 (or a dissimilarity measure of 1). Two networks are similar if they have large numbers of nodes that are similar. To examine this, UCINET uses Quadratic Assignment Procedure (QAP) regression analysis. The correlation coefficient and its respective p-value are then calculated for each pair of networks—see Table 6 and Table 7.

Since the correlation coefficient can assume any value between -1 and 1 , the values in the diagonal of Table 6 are 1 (one) because the correlation between each matrix and itself is perfect. The results of Table 6 must be carefully cross-checked with the p-values in Table 7. To be confident that a strong correlation exists between the structures of two networks, and assuming a 95% confidence interval, the p-values should be lower than 5% . In other words, the correlation coefficient should be statistically significant.

By analyzing the results in Tables 6 and 7, 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 are characterized by having a high volume of piping, static equipment, electric/automation, and utility works. The correlations between the case project networks are not irrelevant: after all, they represent concepts from the same domain (construction of O&G facilities). They are also not high enough to be considered similar, which could render them useless for benchmarking: why comparing projects or attempting to retrieve a previous project case that is highly relevant to a new project if all networks are the same? This question showcases the validity of this research work's central premise: representing unstructured data in the form of a semantic network can enhance the presentation and retrieval of knowledge from these data sources. This claim also highlights the importance of quality checks on networks by a knowledge manager. It is also possible that a more extensive lexicon could help create more distinction in networks.

7. Learning: Blockmodeling

Blockmodeling is a technique that attempts to capture closely connected nodes within a network. While it groups nodes, the blockmodeling aim is different from typical clustering measures. It is a network partitioning approach—not a cluster-finding method. It aims to reduce a large network into a set of small networks. In turn, reducing each small network into a single (super) node results in a simplification of the original network, which is typically large and noisy. So, unlike simplified clustering measures, blockmodeling achieves its goal through qualifying sub-networks (blocks) in contrast to counting node connections. It attempts to rearrange the adjacency matrix iteratively to find smaller cohesive networks (for more on this, see [47]).

Discovering coherent sub-networks within a large network serves two purposes. First, these sub-networks can be promising knowledge constructs. i.e. groups (blocks) of highly related concepts such that they

Table 6

Correlation coefficient of the dissimilarity matrices (using QAP regression) for the three case networks.

	Project A	Project B	Project C
Project A	1	0.516	0.309
Project B	0.516	1	0.333
Project C	0.309	0.333	1

Table 7

P-values of the similarity matrices using QAP regression.

	Project A	Project B	Project C
Project A	0	0.0003	0.0003
Project B	0.0003	0	0.0003
Project C	0.0003	0.0003	0

are entitled to be considered building blocks of our knowledge. The concepts in the blocks and their relationships capture (possibly generic) knowledge that may qualify to be added to an ontology. The second purpose of conducting blockmodeling is to compare blocks across networks to be a more meaningful means to examine network similarity. Here the similarity of the network is not measured based on counting node connections. Instead, it is assessed based on the similarity at the sub-network level, which indirectly refers to the similarity of the knowledge constructs forming the two networks [48].

Fig. 4 shows the blockmodels for two networks. On the left is the blockmodel for Project A. On the right is the blockmodel for the test network (see the Evaluation section). The test network was generated using Green 2.0 through simulated discussions for Project A. Each network is presented as an adjacency matrix. After an iterative process, the matrices are arranged based on the correspondingly connected nodes in each matrix. Each pair is called a block. i.e. a (single) block is a set of two sub-networks that are similar in the two main networks.

It is expected that both networks have a degree of similarity. We present two different approaches to measure to what extent the test and Project A networks are similar. The first method used QAP. The second used a generalized blockmodeling of the two networks. For the QAP analysis, UCINET was used to calculate the correlation factor between both adjacency matrices. The correlation factor found was 0.449 , with a level of significance of 0.0002 (0.02%). The mid-way value indicates a degree of similarity between the two networks' structures—after all, they both resulted from discussing the same project. However, the two networks are also different. This difference, of course, is because two different teams generated the two networks: the same project will be seen and profiled differently by two different teams (a stable characteristic and challenge of dealing with knowledge in project management).

Pajek was the software used to apply a generalized blockmodeling. Iteratively, the software attempts different sizes of blocks. Eventually, the software recommends an optimal number of blocks. The adjacency matrix for Project A shows the nodes arranged in a manner that reflects the optimal solution suggested by the software. The test network nodes were arranged in the same order as Project A nodes to facilitate the comparison. There is a general pattern of similarity in node relationships. Visually, it is clear that the test network is sparser than Project A network. Project A network has 113 relations, which is 36% more than the test network. This observation was expected given that the discussions on Green 2.0 lasted for only 34 days. In the real world, planning and scoping projects take much longer.

To illustrate the concept of blockmodeling, consider the left-bottom corner of each side of Fig. 6. If we set the block size to 3×1 , we can notice a complete match for the following block: type/capacity of equipment, consumption, productivity (vertical dimensions), and rigging (horizontal dimension). This pattern is shown in the innermost rectangle in red. If we set the block size to 3×2 , we can see that a block with the same concepts, including “onsite transportation”, is also a complete match between the two networks (the middle rectangle, shown in green). If we set the block size to 3×3 , adding the concept “scaffolding” to the group will still define a completely similar block in both networks (the outermost rectangle, shown in blue). The more matches with bigger block sizes appear, the more similar the networks are.

Generalized blockmodeling tools provide an index for block similarity—to what extent two networks are similar. First, similarity must be defined. Blocks can be recognized base on several types/levels of

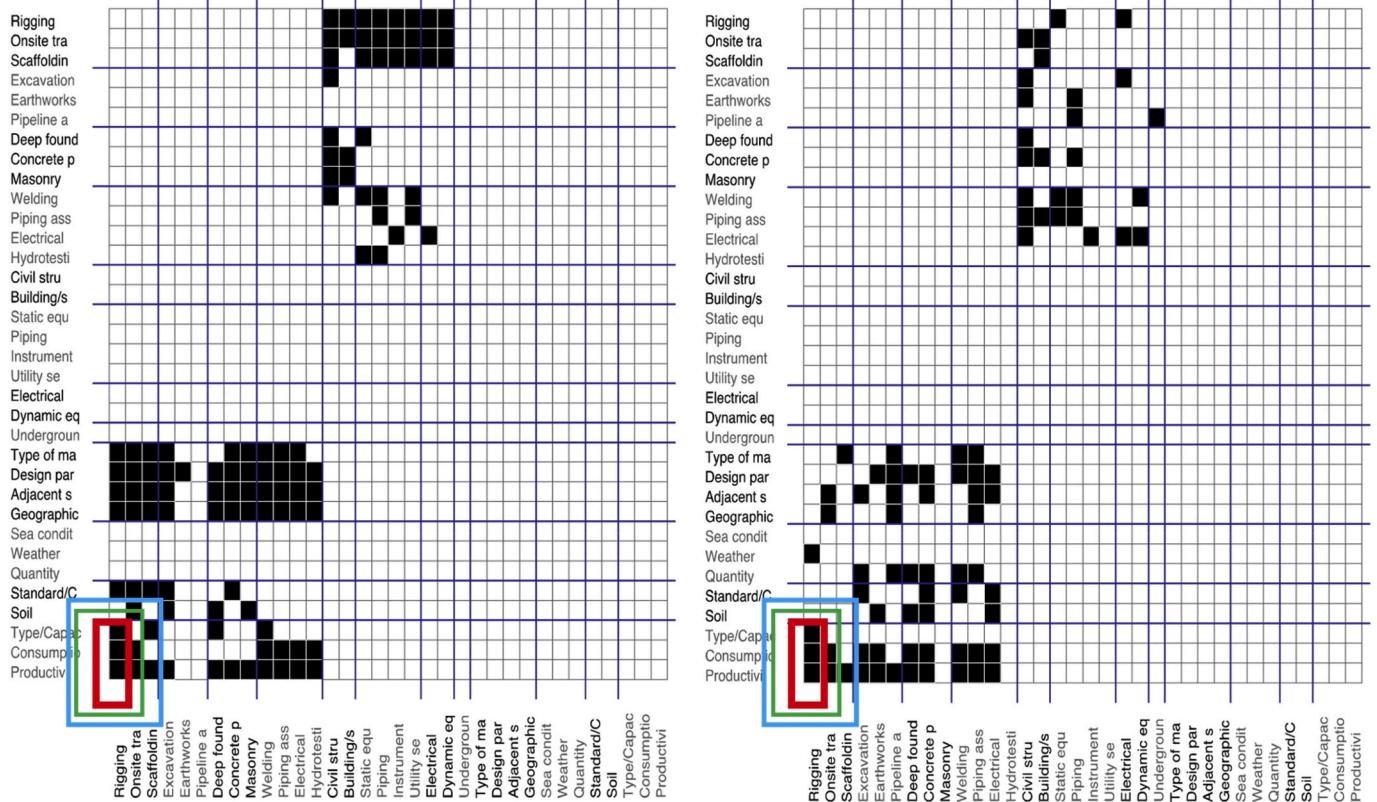


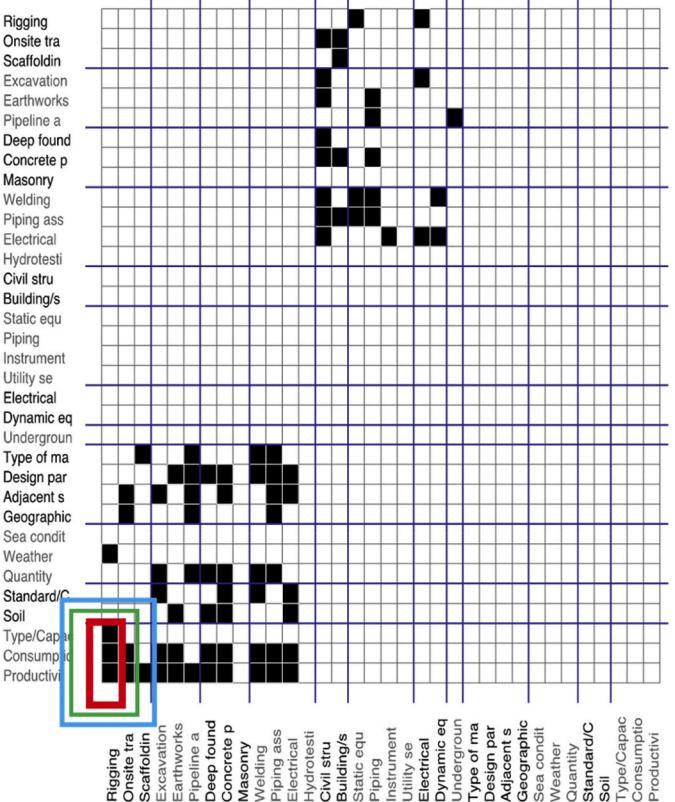
Fig. 4. Project A network (left) and the test network (right) represented as adjacency matrices adjusted based on common blocks.

similarity: complete, null, regular, and row-regular. They are abbreviated as *com*, *null*, *reg*, and *rre*, respectively. Let us also set t to refer to blocks in one network and t' to refer to blocks in the counterpart network. Each block t or t' can assume any of the four types.

Let us also use the counters o and p to count or refer to the locations of blocks in each adjacency matrix—one is for counting on the horizontal direction and the second on the vertical. The index o,p refers to a single block (i.e. a relationship between two nodes). Let us also assume that we set the desired number of blocks to n . The larger the value of n is, the more matches we will find (smaller size blocks). The primary goal here is to compare each block of the $t_{o,p}$ with the corresponding block $t'_{o,p}$. A score is assigned to each block based on its type—for example, a complete (match) block is assigned 1. Eq. (1) presents an index for block analogy (k).

$$k_{o,p} = \frac{r - e}{r} [(t_{o,p} = t'_{o,p}) \vee \{(t_{o,p} = com \vee t_{o,p} = reg) \wedge (t'_{o,p} = reg \vee t'_{o,p} = rre)\}] \quad (1)$$

Where r is the number of possible ties in each block, and e is the error (deviation) of a block $(t_{o,p}, t'_{o,p})$ in the test network from its counterpart (benchmark) block in network A $(t_{o,p}, t_{o,p})$. $k_{o,p}$ assumes any value between 0 and 1 if the expression to the right is true, or 0 otherwise. The expression on the right side of eq. 1 can be true in two cases: if the corresponding blocks in the two networks are the same ($t_{o,p} = t'_{o,p}$); or if the type of a t' block is more “complete” than its corresponding (benchmark) t block. For example, if $t_{o,p} = com$, and $t'_{o,p} = reg$, then $k_{o,p} \neq 0$, since a regular block is a relaxation of a complete block. In summary, complete, regular, and row-regular blocks have a hierarchy that should be taken into account when comparing the blocks. Summing to all blocks to be analyzed, the average block analogy index \bar{k} is given by Eq. (2).



$$\bar{k} = \frac{1}{n} \sum_{o=1}^n \sum_{p=1}^n k_{o,p} \quad (2)$$

In which n is the number of clusters to be analyzed. \bar{k} is a weighted average that accounts for i) how identical the types of blocks of t and t' are, and ii) what is the deviation (the error) from the benchmark block. The average block analogy index \bar{k} can assume any value between 0 and 1; the closer \bar{k} is to 1, the more the match between blocks. The average block analogy index \bar{k} calculated for the two networks was 0.51. Having \bar{k} as 0.51 does not characterize a strong similarity between these networks, but it also indicates that the two networks are positively correlated.

Besides the normal variations due to team formation and human subjectivity, several reasons may have contributed to the dissimilarity between the two networks. The test network was produced through a hypothetical scenario involving actors who were not bidding on the job or planning it. The participants also had mixed, varying levels of expertise in O&G projects, and the experiment with the test network lasted only 34 days. Given this scenario, the test network should be considered an initial network. The network of Project A is a final network, which should have taken much longer to evolve from an initial version.

8. Evaluation

The aim of the evaluation phase of this research work was not to establish a sort of statistical accuracy for the proposed approach in capturing knowledge—for example, a correlation measure of the similarity between networks and the “knowledge contained” in their documents. This task is infeasible. First, this would require the existence of an extensive number of projects and their networks. Furthermore, the fact

that existing documents are to be transferred to the networks manually makes this task too counterproductive. However, it is hoped that using social BIM tools such as Green 2.0, where networks are generated automatically, will make using such a statistical approach more feasible. Second, and more importantly, defining accuracy in this context is not possible. The network is a formalized, objective representation of knowledge. The documents are unstructured and subjective in nature. It is not feasible to develop a numerical accuracy measure for these two heterogeneous forms of knowledge representation.

The best way to judge the value of the proposed approach is to ask experts. The aim of the evaluation focused, then, on two main objectives: asking experts and users how the format and analysis of the networks' structure and contents reflected what they perceived to be the knowledge in their projects; and how suitable and easy was the proposed approach in capturing knowledge convincingly. The evaluation included interviews, a questionnaire, and a forum meeting. We adopted seven criteria to assess the process of collaboration (Table 8), and nine to assess the value and suitability of networks in representing knowledge (Table 9 – see [48,49] for a guideline to develop these questions).

This research also evaluated the suitability of the proposed vocabulary. Can such a simplified lexicon be useful and valuable? It was assessed according to the relevance of the concepts, its completeness, appropriateness of the names, and overall satisfaction. Also, the possibility of describing more complex activities based on the frameworks' thirteen basic construction activities was evaluated. Finally, the Green 2.0 interface was assessed considering three criteria: ease of use, ease of learning, and user satisfaction [50].

9. Contrasting the networks

Before conducting the formal questionnaires and interviews, the first indicator for the proposed approach's value and relevance stems from qualitatively contrasting the networks. The five networks were developed in different ways. The baseline network was generated through a semi-automated approach—answers by the experts surveyed, indirectly, formed the adjacency matrix. Once an adjacency matrix is developed, a network is automatically produced in UCNET. The networks for the three projects were developed through a manual process. The first author developed a final project report, then the two authors debated and developed a network for each project. The participants of each project were asked to check the networks. The test network was developed through the semi-automated process within Green 2.0.

Throughout the research, discussions with participants showed that they found the networks relevant and, in many cases, revealing. More importantly, they pointed out that any of the networks suggested no

Table 8

Criteria used to evaluate the process of collaboration (extracted and adapted from [49]).

Criterion	Definition
Voluntary participation and commitment	Affected or interested stakeholders participate voluntarily and are committed to the process.
Clear ground rules	As the process is initiated, a comprehensive procedural framework is established that includes clear terms of reference, operating procedures, schedule, and protocols.
Equal opportunity and resources	The process allows an equal and balanced opportunity for the effective participation of all interested/affected stakeholders.
Principled negotiation and respect	The process operates according to the conditions of principled negotiation, including mutual respect, trust, and understanding.
Flexible, adaptive, creative	Flexibility is designed into the process to allow for adaptation and creativity in problem-solving.
High-quality information	The process incorporates high-quality information into decision making.
Effective process management	The collaborative process is managed and coordinated effectively and in a neutral manner.

Table 9

Criteria used to assess the value/suitability of networks in representing knowledge (Extracted and adapted from [49]).

Criterion	Definition
Agreement	The process reaches an agreement (a network) accepted by all stakeholders.
Perceived as successful	The process and network are perceived as successful by stakeholders.
Conflict reduced	The network (and process) reduces conflict.
Superior to alternative	The network (and process) is perceived as being superior to the alternative.
Innovation and creativity	The process/network produces innovative ideas and outcomes.
Knowledge, understanding, and skills	Stakeholders gained knowledge, understanding, and skills by participating in the collaborative process.
Information	The network (and process) produces improved information and analyses through joint fact-finding that stakeholders understand and accept as accurate.
Common interest	The outcomes are regarded as meeting the common good or larger interest of participants, and not just the interests of stakeholders involved.
Understanding and support of CPs	The process results in an increased understanding of, and participants support for, collaborative processes/collaborative stakeholder groups.

unrealistic or strange aspects. The positive reception of networks developed through different approaches can mean that the basic idea of a network representing knowledge is feasible and relevant. Of particular interest is the comparison between Project A network and the test network. The two networks were developed for the same project through two different approaches. Despite the limitations of the process of generating the test network, both networks were close enough to reflect the scope of the project; and different enough to capture the differences in the teams' composition and the analysis time.

Interestingly, this result showcases the main hypothesis of this research work. Different teams will produce different knowledge, even for similar projects. Hence, retrieving knowledge (networks) from previous projects can be informative to a team working on new projects. Recognizing that the team of the test network lacked experience, the retrieval of older, similar cases can be a source of training for new staff. Finally, recognizing that the test network is akin to an initial network and that Project A network is a final project network justifies the idea that comparing interim project networks can be valuable. i.e. the test networks can be seen as an initial version of Project A network. A knowledge manager can use these differences to study the evolution of issues in a project.

10. User evaluation

After finishing the debate on the simulated project A on Green 2.0, users were asked to respond to an online questionnaire and participate in a focus group, an unstructured group interview. The back and forth open discussions allow the researcher to probe and clarify their views [51]. A focus group is a meaningful exploratory method to assess new concepts, system or product usability, innovative processes, and marketing campaigns, amongst others [52].

In both evaluation methods, the participants assessed the following aspects of the proposed approach.

- The lexicon;
- The process of collaboration;
- The value of the outcome of the simulated project case (through Green 2.0).

The online questionnaire was submitted to the participants two days before the end of the simulated project run. To minimize bias, the participants responded anonymously to 67 questions distributed according to the parts shown in Table 10.

Table 10

Criteria included in the online questionnaire and a corresponding number of questions.

Parts	Number of questions
Suitability of the lexicon	10
Process of collaboration	14
Outcome (networks)	18
Green 2.0 as an interface	18
Demographics and the test experience	7
Total	67

The ten questions concerning the suitability of tags targeted the lexicon's three dimensions (categories), appropriateness of concept names, and overall satisfaction. Also, participants were asked about its adaptability for changes to build more complex ones. The aim here was not to validate whether the proposed lexicon is a complete taxonomy of the concepts, which was not the objective for developing it. Instead, 1) to evaluate if it was good enough to create a common vocabulary, and 2) to examine if the small-sized lexicon was restrictive to users. Negative answers would cast doubts about the relevance of the test network.

The questions regarding the collaboration process were based on the criteria in [Table 8](#) and [Table 9](#). Finally, 18 questions covered three evaluation criteria: ease of use, ease of learning, and overall satisfaction with the system's interface. For the questions regarding the tags and the lexicon, a 5-point Likert scale was used (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree). For the remaining assessment topics, a 4-point Likert scale was more appropriate (strongly agree, agree, disagree, or strongly disagree). At the end of each part, an open discussion encouraged users to provide general evaluations and suggestions.

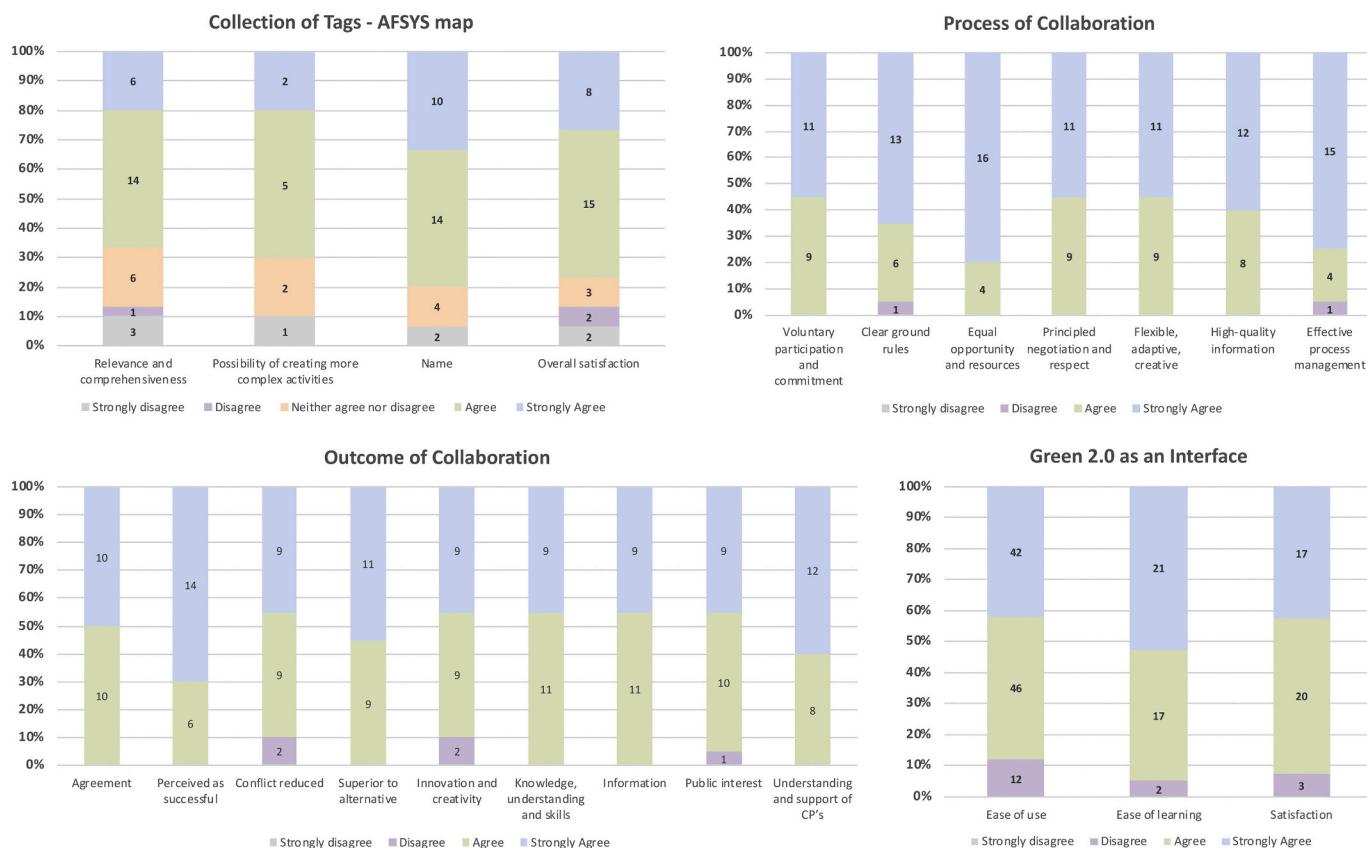
[Fig. 5](#) (top left panel) shows the results for the lexicon questions. The combined score for "agree" and "strongly agree" were as follows:

relevance and comprehensiveness of the lexicon: 66.7%; possibility of using the lexicon for creating additional or more complex activities: 90%; semantics (appropriateness) of activities naming: 70%. The combined "agree and "strongly agree" for the overall satisfaction with the lexicon was 76.7%. It should be noted that "relevance and comprehensiveness" received "strongly disagree" or "disagree" in 13.3% of answers; and received "neither agree nor disagree" in 20% of the answers. However, the vast majority (90%) recognized that creating more complex activities using the 13 basic concepts is viable. This result means that while the users may have disagreed on the lexicon's comprehensiveness, they agreed on its relevance.

Regarding the process of collaboration ([Fig. 5](#), top right panel) and satisfaction with its outcomes ([Fig. 5](#), bottom left panel), "agree" or "strongly agree" was the answer in 90% of the answers. Before that, the participants were asked to evaluate the Green 2.0 interface and its usability. This evaluation aimed at making sure the results were not influenced by the interface (see [Fig. 5](#), bottom right corner). About 90% of the responses were positive. Finally, two questions investigated the level of understanding of the scope of project A before and after the test. The participants were asked to answer using a rating scale from 0 to 10. The users indicated that their level of understanding was 6.6 before the beginning of the test and 8.2 after. This result also suggests that the test was useful in exchanging knowledge.

11. Focus group

The objective of the focus group was to collect the participants' feedback in a more unstructured way. The participants were divided into two groups, and a focus group meeting was held with each group separately. Each meeting lasted two hours and started with an introduction to its purpose. A handout with the main questions and evaluation criteria was distributed to participants. A presentation about the



[Fig. 5](#). Results of the questionnaire.

proposed approach and the evaluation results followed. The handout also contained the four dimensions of assessment and a summary of the criteria. The participants were encouraged to discuss their evaluation of all criteria without input from the moderator. Then the participants were asked to compile their findings in a short report. For that, the participants engaged in a final discussion and some negotiation to reconcile differences in views and synthesize their main findings.

One of the main discussed issues was the ease of the system. Participants agreed that the Green 2.0 platform is intuitive. They also agreed that placing comments directly to BIM is much more efficient than the traditional way of sending e-mails or having meetings. It was beneficial to draw on the expertise of others when a participant had a question. Also, reviewing the general discussion topics and their trends was informative. One contentious issue was how many base concepts should be used in the lexicon: having too many tags increases the user's burden, and having too few tags may restrict the user in selecting the most appropriate tags in their comments. To an extent, this topic is the Achilles' heel for any modelling exercise: scope and depth—how many concepts to include and for what specific purpose? This issue is immaterial to the evaluation of the proposed approach as the base lexicon's size is a decision that is to be taken based on the conditions and preferences of the team using it—they can update the lexicon to include any level of granularity they prefer.

12. Conclusions

Managing unstructured data is a challenge. It is hard for information management to process and retrieve previous data (documents) in a formalized way. More importantly, for knowledge management, formalizing its rich tacit knowledge is also challenging. This research showcased the use of semantic networks to represent the contents of documents formally. The nodes in the network are the main concepts in documents, and the links are a representation of a semantic relationship between them. Project documents can be transferred into a semantic network through a manual process—either by a project team, by a dedicated knowledge manager, or through collaboration between them. While the lack of automation in this regard is time-consuming, research into knowledge representation and retrieval has shown that the value of human interpretation is still unmatched. The proposed approach presented one option for an automated generation of networks: a BIM-based interaction platform. On the one hand, it makes chats between project stakeholders easier; and, at the same time, provides means to capture this new source of unstructured data. On the other hand, it automatically generates a semantic network of the chats.

Studying the knowledge constructs (groups of concepts) in a document can be conducted using several network analysis measures, such as centrality, diameter, length, and clustering coefficients. They can reveal, through different metrics, various aspects of influential concepts and their relationships. Cross-network comparisons can help contrast two networks. This approach can rely on measures such as dissimilarity indexes and blockmodeling. Studying the contrast between interim project networks can help explore the evolution of issues and their importance. Contrasting an interim project network to a database of past projects' networks can help retrieve documents from projects with conditions that best match the ongoing project.

In addition to using it for contrasting networks, blockmodeling is a promising technique for capturing knowledge constructs. It seeks to find blocks of linked concepts that are similar in two networks. Observing the existence of a block repeatedly in different projects is a strong indication that it represents a generic knowledge construct. This process can be beneficial in building or updating an ontology.

Following the open-system theory (or the no-model) approach and the field theory, this proposition serves co-creation and evolutionary knowledge production purposes. It allows collaboration where different knowledge is exchanged; simultaneously, it provides stakeholders with a formalized view of what they have been discussing through the semantic

network of their comments. This idea enables them to 1) see a possible externalization of their tacit knowledge and 2) promote co-learning and collaborative evaluation of knowledge. This way, knowledge is extracted and formally represented bottom-up and can balance the dominating top-down approach of ontologies, making some of them overly theoretical. Here, knowledge reflects what happened in the field, as seen and screened (negotiated) by practitioners.

Social BIM has significant advantages, especially as stakeholders move towards a culture where chatting and texting have become essential communication modes. It is not expected that social BIM will replace all other modes. Therefore, organizations need to consider the means to create an easy flow of information and integration across different communication media. As with any other system, the lexicon's granularity is another typical issue with any semantic modelling. No one solution can fit all organizations and contexts. However, one main advantage of tracking chats is that automated topic modelling systems can learn frequent keywords and adapt the lexicon to the ongoing discussion. Finally, social BIM may experience typical social media problems—particularly the behavioural and sociological issues that arise from forming any community.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Y. Rezgui, S. Boddy, M. Wetherill, G. Cooper, Past, present and future of information and knowledge sharing in the construction industry: towards semantic service-based e-construction? *CAD Comput. Aided Des.* 43 (2011) 502–515, <https://doi.org/10.1016/j.cad.2009.06.005>.
- [2] P. Pauwels, S. Zhang, Y.-C. Lee, Semantic web technologies in AEC industry: a literature overview, *Autom. Constr.* 73 (2017) 145–165, <https://doi.org/10.1016/j.autcon.2016.10.003>.
- [3] M. Grobelnik, D. Mladenic, B. Fortuna, Semantic Technology for Capturing Communication inside an organization, *IEEE Internet Comput.* 13 (2009) 59–67, <https://doi.org/10.1109/MIC.2009.88>.
- [4] S. De Deyne, D.J. Navarro, A. Perfors, G. Storms, Structure at every scale: a semantic network account of the similarities between unrelated concepts, *J. Exp. Psychol. Gen.* 145 (2016) 1228–1254, <https://doi.org/10.1037/xge0000192>.
- [5] V. Wickhorst, Measuring the ROI on knowledge management systems, *Perform. Improv. Q.* 15 (2008) 96–127, <https://doi.org/10.1111/j.1937-8327.2002.tb00252.x>.
- [6] P. Checkland, Model validation in soft systems practice, *Syst. Res.* 12 (2007) 47–54, <https://doi.org/10.1002/sres.3850120108>.
- [7] G.E.P. Box, Science and statistics, *J. Am. Stat. Assoc.* 71 (1976) 791, <https://doi.org/10.2307/2286841>.
- [8] S.M. Lee, D.L. Olson, S. Trimis, Co-innovation: convergenomics, collaboration, and co-creation for organizational values, *Manag. Decis.* 50 (2012) 817–831, <https://doi.org/10.1108/00251741211227528>.
- [9] T.G. Cummings, E.S. Molloy, *Improving Productivity and the Quality of Work Life*, 1st Ed, Praeger, New York, 1977.
- [10] T. El-Diraby, T. Krijnen, M. Papagelis, BIM-based collaborative design and socio-technical analytics of green buildings, *Autom. Constr.* 82 (2017) 59–74, <https://doi.org/10.1016/j.autcon.2017.06.004>.
- [11] P. Teicholz, White Paper on AEC Practice and Research Issues for the Future, in: Berkeley-Stanford Work, Defining a Res. Agenda AEC Process Prod. Dev, Berkeley-Stanford, Stanford, 1999, pp. 1–3. <http://faculty.ce.berkeley.edu/tommemein/CEMworkshop/Teicholz.pdf>.
- [12] N.A. Kartam, Making effective use of construction lessons learned in project life cycle, *J. Constr. Eng. Manag.* 122 (1996) 14–21, [https://doi.org/10.1061/\(ASCE\)0733-9364\(1996\)122:1\(14\)](https://doi.org/10.1061/(ASCE)0733-9364(1996)122:1(14)).
- [13] M. Marlin, Implementing an Effective Lessons Learned Process in a Global Project Environment, in: UTD 2nd Annu. Proj. Manag. Symp. Proc. –Dallas, Texas, Dallas, 2008, pp. 1–6.
- [14] S.J. Simoff, M. Lou Maher, Ontology-Based Multimedia Data Mining for Design Information Retrieval, in: Congr. Comput. Civ. Eng. Proc., Cambridge, 1998, pp. 1–11.
- [15] F.B. Aalami, M.A. Fischer, J.C. Kunz, AEC 4D Production Model: Definition and Automated Generation. <https://purl.stanford.edu/yt829pn1767>, 1998.
- [16] L. Soibelman, F. Peña-Mora, Distributed multi-reasoning mechanism to support conceptual structural design, *J. Struct. Eng.* 126 (2000) 733–742, [https://doi.org/10.1061/\(ASCE\)0733-9445\(2000\)126:6\(733\)](https://doi.org/10.1061/(ASCE)0733-9445(2000)126:6(733)).

- [17] C.M. Eastman, Building Product Models: Computer Environments, Supporting Design and Construction, CRC Press, 2018, <https://doi.org/10.1201/9781315138671>.
- [18] C.H. Caldas, L. Soibelman, J. Han, Automated classification of construction project documents, *J. Comput. Civ. Eng.* 16 (2002) 234–243, [https://doi.org/10.1061/\(ASCE\)0887-3801\(2002\)16:4\(234\)](https://doi.org/10.1061/(ASCE)0887-3801(2002)16:4(234)).
- [19] J. Zhang, N. El-Gohary, Extraction of Construction Regulatory Requirements from Textual Documents Using Natural Language Processing Techniques, in: *Comput., Civ. Eng.*, American Society of Civil Engineers, Reston, VA, 2012, pp. 453–460, <https://doi.org/10.1061/9780784412343.0057>.
- [20] P. Zhou, N. El-Gohary, Ontology-based multilabel text classification of construction regulatory documents, *J. Comput. Civ. Eng.* 30 (2016), 04015058, [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000530](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000530).
- [21] Ž. Turk, Phenomenological foundations of conceptual product modelling in architecture, engineering and construction, *Artif. Intell. Eng.* 15 (2001) 83–92, [https://doi.org/10.1016/S0954-1810\(01\)00008-5](https://doi.org/10.1016/S0954-1810(01)00008-5).
- [22] M.N. Bakht, T.E. El-Diraby, Synthesis of decision-making research in construction, *J. Constr. Eng. Manag.* 141 (2015), 04015027, [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000984](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000984).
- [23] M. Newman, Networks: An Introduction, 1st ed., Oxford University Press, New York City, NY, 2010 <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>.
- [24] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393 (1998) 440–442, <https://doi.org/10.1038/30918>.
- [25] S.P. Borgatti, M.G. Everett, Regular blockmodels of multiway, multimode matrices, *Soc. Networks* 14 (1992) 91–120, [https://doi.org/10.1016/0378-8733\(92\)90015-Y](https://doi.org/10.1016/0378-8733(92)90015-Y).
- [26] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Application*, 1st Editio, Cambridge University Press, New York, 1994.
- [27] P. Doreian, V. Batagelj, A. Ferligoj, *Generalized Blockmodeling*, 1st Editio, Cambridge University Press, New York City, NY, 2005.
- [28] S.P. Borgatti, M.G. Everett, Two algorithms for computing regular equivalence, *Soc. Networks* 15 (1993) 361–376, [https://doi.org/10.1016/0378-8733\(93\)90012-A](https://doi.org/10.1016/0378-8733(93)90012-A).
- [29] V. Batagelj, A. Ferligoj, P. Doreian, Direct and indirect methods for structural equivalence, *Soc. Networks* 14 (1992) 63–90, [https://doi.org/10.1016/0378-8733\(92\)90014-X](https://doi.org/10.1016/0378-8733(92)90014-X).
- [30] J. Chan, S. Lam, C. Hayes, Generalised blockmodelling of social and relational networks using evolutionary computing, *Soc. Netw. Anal. Min.* 4 (2014) 1–24, <https://doi.org/10.1007/s13278-014-0155-y>.
- [31] E.S. Sagor, D.R. Becker, The effect of data collection technique on estimated landowner personal network attributes, *Small-Scale For.* 13 (2014) 299–317, <https://doi.org/10.1007/s11842-013-9255-1>.
- [32] A. Singhal, Introducing the Knowledge Graph: Things, Not Strings, Google. <https://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html>, 2012.
- [33] A. Fader, S. Soderland, O. Etzioni, Identifying relations for open information extraction, in: Proc. Conf., Association for Computational Linguistics, Edinburgh, Scotland, 2011, pp. 1535–1545, <https://doi.org/10.1234/12345678>.
- [34] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D.S. Weld, A. Yates, Web-Scale Information Extraction in KnowItAll (Preliminary Results), in: *WWW'04 Proc. 13th Int. Conf. World Wide Web*, New York City, NY, 2004, pp. 100–110, <https://doi.org/10.1145/988672.988687>.
- [35] E. Reiter, R. Dale, *Building Natural Language Generation Systems*, 1s edition, Cambridge University Press, Cambridge, 2000.
- [36] M. Nik-Bakht, T.E. El-Diraby, Sus-tweet-ability: exposing public community's perspective on sustainability of urban infrastructure through online social media, *Int. J. Hum. Comput. Stud.* 89 (2016) 54–72, <https://doi.org/10.1016/j.ijhcs.2015.11.002>.
- [37] Z.A. Adamu, S. Emmitt, R. Soetanto, Social BIM: co-creation with shared situational awareness, *J. Inf. Technol. Constr.* 20 (2015) 230–252, <https://www.itcon.org/paper/2015/16>.
- [38] J. Zhang, T.E. El-Diraby, SSWP: a social semantic web portal for effective communication in construction, *J. Comput.* 4 (2009) 330–337, <https://doi.org/10.4304/jcp.4.4.330-337>.
- [39] E. Trist, H. Murray, *The Social Engagement of Social Science, Volume 2: A Tavistock Anthology - the Socio-Technical Perspective*, University of Pennsylvania Press, Philadelphia, 1990.
- [40] B. Burnes, B. Cooke, Kurt Lewin's Field Theory: A Review and Re-evaluation, *Int. J. Manag. Rev.* (2012), <https://doi.org/10.1111/j.1468-2370.2012.00348.x> n/a-n/a.
- [41] William Pasmore, Action research in the workplace: The socio-technical perspective, in: *Handb. Action Res.*, Sage Publications, 2006, pp. 38–48.
- [42] P. Yung, K.C. Lam, C. Yu, An audit of life cycle energy analyses of buildings, *Habitat Int.* 39 (2013) 43–54, <https://doi.org/10.1016/j.habitatint.2012.10.003>.
- [43] M.M. Bilec, R.J. Ries, H.S. Matthews, Life-cycle assessment modeling of construction processes for buildings, *J. Infrastruct. Syst.* 16 (2010) 199–205, [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000022](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000022).
- [44] A.A. Guggemos, A. Horvath, Decision-support tool for assessing the environmental effects of constructing commercial buildings, *J. Archit. Eng.* 12 (2006) 187–195, [https://doi.org/10.1061/\(ASCE\)1076-0431\(2006\)12:4\(187\)](https://doi.org/10.1061/(ASCE)1076-0431(2006)12:4(187).
- [45] M. Buyle, J. Braet, A. Audenaert, Life cycle assessment in the construction sector: a review, *Renew. Sust. Energ. Rev.* 26 (2013) 379–388, <https://doi.org/10.1016/j.rser.2013.05.001>.
- [46] G. Moröl, S. Kim, Network text analysis and social network analysis in investigating complex governance networks: applications of AutoMap and ORA, in: *Handb. Res. Methods Complex. Sci.*, Edward Elgar Publishing, 2018, pp. 199–220, <https://doi.org/10.4337/9781785364426.00021>.
- [47] P. Doreian, An intuitive introduction to blockmodeling with examples, *Bull. Sociol. Methodol. Méthodologie Sociol.* 61 (1999) 5–34, <https://doi.org/10.1177/075910639906100103>.
- [48] 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, <https://doi.org/10.1016/j.aei.2019.01.003>.
- [49] N. Cradock-Henry, Evaluating a collaborative process, *Policy Br* 2 (2013), <https://doi.org/10.13140/RG.2.2.27571.43043>. PG1–4.
- [50] A.M. Lund, Measuring usability with the USE questionnaire, *Soc. Tech. Commun.* 8 (2001) 3–6, <https://doi.org/10.1177/1078087402250360>.
- [51] H. Edmunds, *The Focus Group Research Handbook*, 1st Editio, Emerald Group Publishing Limited, Chicago, IL, 1999, <https://doi.org/10.1108/bl.1999.12.3.46.1>.
- [52] B. Nagle, N. Williams, *Methodology Brief : Introduction to Focus Groups*. <http://www.uncfsp.org/projects/userfiles/File/FocusGroupBrief.pdf>, 2011.