Advanced Engineering Informatics 39 (2019) 214–226

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Using blockmodeling for capturing knowledge: The case of energy analysis in the construction phase of oil and gas facilities

Rodrigo Rodrigues Aragao*, Tamer E. El-Diraby

Department of Civil Engineering, University of Toronto, Canada

ARTICLE INFO

Keywords:
Knowledge retrieval and representation
Network analysis
Blockmodeling
Construction
Energy assessment
Oil and gas

ABSTRACT

In this paper, blockmodeling, a network analysis clustering approach, is used to study and capture clusters of concepts. The semantics networks are a former stage of the extracted, formalized knowledge from unstructured data (text). As a sample domain of the application of the proposed approach, the networks are focused on concepts related to planning for energy management during the construction phase. Text describing the status or lessons learned from projects is transferred to semantic networks. A benchmark concept network was generated based on surveying experts. Blockmodeling algorithms were used to create a set of concept blocks: clusters of related concepts that capture some of the knowledge of a generic scenario. Through interviews with staff from three projects, we developed 3 case studies (text) to capture the conditions and knowledge gained in these projects. A concept network of main concepts was extracted for each case. Also, blockmodels from these networks were also extracted. To facilitate the comparison, an average block analogy index $\bar{k}$ was introduced. The smaller $\bar{k}$ is, the more dissimilar the studied blocks are, and the more unique and unusual are the characteristics of the project at hand. By contrasting the blocks of the case projects against each other and against the benchmark network, we identified unique knowledge constructs (concept clusters) in the three projects. This can be beneficial in capturing project-specific knowledge; contrasting project conditions and knowledge concepts; and supporting a frequent upgrade of the benchmark concept network or a formal ontology.

1. Introduction

Project exchanges, such as documents, e-mails, and minutes of meetings encapsulate valuable knowledge. To capture some of this knowledge, companies typically develop a summary report or a project case study to support future projects. To formalize the semantics of such unstructured data, researchers have used two major approaches. First approach uses topic modeling, where concepts within a corpus are clustered into topics based on their co-location and frequency of appearance within documents. A key problem with this approach is that the topics do not necessarily make sense or capture proper semantics of the group of words [1]. However, topic modeling has a dynamic nature, extracts the clusters bottom-up, and is context sensitive (reflects the contents of the text corpus). The second approach uses an ontology, which is a structured, formalized set of concepts with specific relations and axioms. Ontology provides a stable standard against which semantics can be evaluated. However, the static nature of ontology and its limited ability to adapt to the specific context of a corpus may limit its value [C].

We propose a semi-automated approach to help find patterns or concept clusters within text. It can possibly be used to enrich ontology development and to enhance topic modeling efficiency. In the proposed approach, text from project reports is transferred into a semantic network. The network is developed based on a map of concepts—loosely connected terms that are used to capture general topics of project knowledge. Blockmodeling is used to discover patterns of concepts.

Blockmodeling is part of network science. It is an iterative procedure that aims at grouping nodes in a network by analyzing equivalencies between them [2]. Discovering blocks can help reduce a large network into a smaller one where each block becomes a node in the new network. More importantly, when similar blocks exist in two networks, we can expect a certain degree of equivalency between them. For example, a recent study used blockmodeling to study the patterns of association between judges on USA Supreme Court and its relationship to the nature of the case being considered and the final decision [3]. The process can help litigants form a preliminary prediction about who are the potential judges that are most likely to agree with their arguments.

Our assumptions is that every project is unique because each has a specific scope and, more importantly, a particular context. However,

* Corresponding author.
E-mail addresses: r.aragao@petrobras.com.br, r.aragao@alum.utoronto.ca (R.R. Aragao), tamer@ecf.utoronto.ca (T.E. El-Diraby).

https://doi.org/10.1016/j.aei.2019.01.003
Received 23 February 2018; Received in revised form 16 January 2019; Accepted 17 January 2019
Available online 24 January 2019
1474-0346/

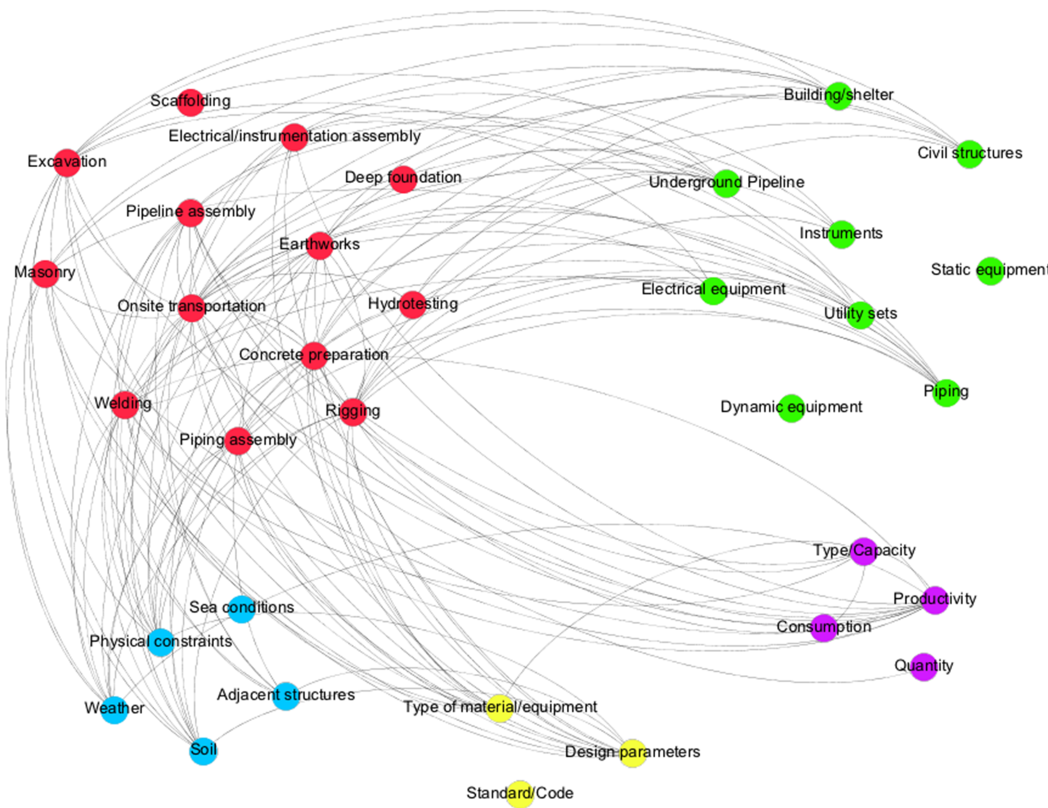

Fig. 1. The concept network for a natural gas pipeline construction project (Project C).

irrespective of their possible sheer differences, there is a common set of concepts that exist in almost all projects. These concepts formulate a mosaic of concepts covering technical and non-technical aspects of projects. Concept blocks can be used to profile the project—akin to developing a semantic signature of a project. This can be used to study how unique a project is based on contrasting it blockmodels to those of a base/benchmark set of blocks (possibly parts of an ontology). The blockmodels can be used to update the ontology or guide the refinement of automatically discovered topics. What makes a project unique is the subset of concepts that it contains and their interrelationships: how are they configured or blocked together. Arrangements of these concepts represent some basic knowledge constructs. Capturing many of these can show the diversity of concept relationship dynamics (possibilities of concept co-existence and interlinkage), which can be very helpful in guiding ontology development.

In the next sections, we first detail the scope of our work and its contribution to the processing of unstructured data in a manner that can capture project-specific knowledge constructs. The following section discusses the methodology of our work, which is meant to detail the needed steps and our approach in handling the analysis. The next section reviews related works, especially in the areas of semantic networks, capturing and using project knowledge using experience-based methods, the complexity of energy analysis for the construction phase, and the fundaments of blockmodeling. Then the three cases and their concept networks are presented. Next, the details of the blockmodeling analysis and the comparison between networks are described in the discussion section. We conclude with the lessons learned and implications of our work.

It is important to point out here that we are focused only on knowledge representation—especially, the extraction of formalized knowledge from text. So, this work is not about developing or selecting an energy analysis process/model, nor do we assume one. We are more interested in capturing concept patterns from stakeholders/decision makers' discussions.

The process of decision making and planning for energy in the construction phase of oil and gas is complex. This is due to the diversity and subjectivity of analysis concepts. Different configurations of the construction method, resources used and the sequencing of each activity can have an impact on the energy used. For example, the pre-assembly of pipes can typically save energy during construction; use of sensor-enabled excavators can increase accuracy and reduce re-work. Modular specifications and simplification of designs can have a major impact on the energy used. The nature of the boundary conditions to the project (such as access and soil conditions) can also limit the energy-saving options. Different stakeholders contribute to defining the activities and their scope, including, the designers, the contractor, subcontractors and regulators. They negotiate several configurations for the design; options for construction tasks and for costing risks and resources. Their debates cover several topics: safety, budget, schedule, environmental impacts, and energy consumption. Our main motivation is that such debates and rebuttals encapsulate context-specific, multidisciplinary knowledge that can revisited in future projects [4].

Our aim is to capture and represent some of the knowledge contained in the input/views of project stakeholders. This is part of a larger domain of knowledge extraction from unstructured data (text, in our case). We decided to use blockmodeling because, first, it can deal with clusters of concepts in a network. Viewing knowledge as a concept network is advantageous for capturing the semantics. By linking concepts to each other, we can explore a deeper understanding of the "meaning" of the text. Second, within network analysis practices/tools, blockmodeling goes beyond simplified concept association. Iteratively, it helps examine different concept blocks (clusters). This approach triangulates a group of concepts, and produces more coherent, relevant topics: they are not connected only by a frequent word association. In fact, a block can be considered as a mini-ontology. This "ontology" has one major advantage: by being driven bottom-up, the block reflects knowledge that is specific to the project at hand. Finally, across-projects, we can use the blocks to study patterns of knowledge representations: variations in blocks from project to project or from project to a benchmark can help understand the nature and stability of concept relations/clustering.

2. Methodology

Given the iterative and interpretive nature of this process, the approach presented is semi-automated, in which an expert has to manipulate and iterate with blockmodeling tools to examine and refine the most suitable blocks of concepts. Our objective is not to develop a formal model or algorithm for blockmodeling in project unstructured data. Rather, we aim at conducting a case study to examine and showcase the potential applications and benefits of blockmodeling in this domain. As such, we conducted three case studies of three oil and gas projects in Brazil, with a specific focus on the planning and analysis of energy consumption during the construction phase. For each, we transferred the text of the case studies into a semantic network (see example in Fig. 1). For additional information about the concept map and the methodology to represent the case studies as a concept network, please refer to [5]. We developed a map of concepts that can be common to most projects. It included three categories of concepts:

- Construction activities: a set of typical construction tasks (red nodes in Fig. 1).
- Systems: the main (physical) components of the project—what actually is being built (green nodes in Fig. 1).
- Factors: these are the boundary conditions that can range from the physical constraints (blue nodes) to design features (yellow nodes) and resources (purple nodes).

We then contrasted the blockmodels of each case to those of a benchmark network. It important to note that the aim is not to develop an ontology *per se* or a standardized set of concepts. This fact highlights the primary philosophy of the proposed approach: finding patterns of blocks can help enrich an existing ontology or topic modeling practices. The proposed map of concepts acts only as the initial seed for possible future ontologies.

The main steps of the methodology are as follows:

Literature review: through synthesizing previous related work, we present a base concept map: concepts that can exist or be used to describe knowledge in a general project. The matrix serves as simplified taxonomy of concepts associated with the energy use in construction.

Case studies: in a previous study [5], we develop three case studies of past oil and gas projects in Brazil. In-depth semi-structured interviews of team members developed a text summary of the project. The participants responded to open-ended questions about the challenges, best practices, opportunities, and risks that affects or could affect the energy use during the construction phase. A profile of responses was generated for each interview. In each profile, the research team identifies keywords or expressions that semantically resembles the concepts from the matrix. A relation or a link is established whenever there is a semantic association between two concepts that affects the energy use in the construction phase. These relations are transferred to an adjacency matrix, and a network of concepts is represented for each case study, such as the one in Fig. 1.

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

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

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

Blockmodeling: Following the steps above, a generalized blockmodeling approach was used to create a benchmark blockmodel, as describe below:

- A set of concept clusters centered around construction activities was created to act as an initial seed. These were developed based on the typical functions for each activity in a general project. Of course, a user of the system can select different clusters based on the context of use.
- As an input for the blockmodeling, we define the number of clusters and the type of blocks allowed to form during the optimization. Based on the clusters of the activities (above), the algorithm output is a blockmodel that attempts to reorder the remaining concepts in such a way that it fulfills the pre-defined clusters and these requirements.
- The adjacency matrix of the case networks are rearranged to fit the blockmodel and contrasted to the benchmark network. Our interest is to study logical blocks in the case networks that are significantly different from those of the benchmark network. Blocks that do not match the base blocks are interpreted in light of the context of the case studies. Blocks with unique, new concept arrangements are the ones that capture project-specific knowledge (concept inter-relationships).
- We introduce a block analogy index that quantitatively assesses the similarities in the networks (benchmark against the case networks) and the block-to-block similarities. The less similar the case network is to the benchmark, the more unique, context-specific is the knowledge contained in the network.

3. Literature review

Construction projects are diverse with ever-changing challenges. This dynamic context of construction projects makes them unique and difficult to plan—especially in regards to energy consumption during the construction phase. Estimates for energy consumption during the construction stage vary markedly. Between 3% and 15% of the energy consumed to manufacture project materials are consumed during construction, with some earlier research claiming as high as 30% [6]. However, more recent research provided more reliable estimates, with a range of about 10% of the embodied energy in the facility [7–10].

There are limitations concerning works that can help decision-makers manage energy consumption or understand its interplay with other past project. This scenario is where experience-based methods and case-based reasoning (CBR) can be helpful. However, capturing, mapping and reusing lessons learned from the case text is not efficient in searching/retrieval of previous cases [11,12]. Promising knowledge representation and retrieval tools are available, such as Google's Knowledge Graph [13], OPEN IE [14], and KnowItAll [15]. However, using automated topic extraction and modeling in such subjective

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. 2. The concept map (taxonomy).

domain may lead to vagueness and inaccuracy [16]. The alternative approach is to develop and use an ontology. However, they are laborious, difficult to maintain and static, with limited abilities to accommodate context [17,18].

The knowledge representation of free texts as a network of concepts is not new. Semantic networks map entities/concepts and define the interrelationship between them [19]. They have been developed to support NLP through representing sentences of free texts [20–22]. Network analysis relies on graph theory and mathematical indicators to investigate prominent nodes, to study equivalent positions and to detect sub-groups or communities [23]. In the construction industry, studies have demonstrated the benefits of network analysis in understanding actor relationships [24]; community discussions about a project on social media [25,26]; communication barriers in construction work teams [27,28]; and the role of communication network in assuring health and safety [29]. Networks of decision making data in highway construction were used to assesses how well an organization uses collected data in their processes [30].

In network analysis, clustering refers to identifying cohesive groups of nodes. Blockmodeling is a clustering method that attempts to group the nodes based on a similarity measure [31], such as structural and regular equivalences. Two nodes are structurally equivalent if they are connected to the same (exact) nodes. Regular equivalence is a relaxation of the structural equivalence, where nodes need only to share connections with other nodes that are themselves equivalent [32]. In social network, equivalent nodes are said to have overlapping roles and therefore their positions are interchangeable. In our domain, equivalent concepts are thought of as holding a collective influence on the network. Blockmodeling captures a contextual knowledge that describes the actions of node groups. It should be analyzed as much as any node-level analysis.

In blockmodeling, the set of relations of a network is presented by an adjacency matrix [33]. The rationale behind this method is to organize the network into specific patterns of relations, called blocks. A block is formed by clusters of nodes with a common pattern of ties [2]. The perfect set of block types that the user expects to find is called the blockmodel, and it is an assumption defined beforehand based on prior information about the network. As such, blockmodeling is a permutation algorithm that attempts to fit the blockmodel to the dataset of the network [2].

Many clustering algorithms can be used for blockmodeling. In the indirect method, the algorithm attempts to optimize a (dis)similarity measure [34–36]. Hierarchical clustering is typically used for indirect blockmodeling, where nodes are grouped based on desired measures, such as the Euclidean distance. Examples of indirect algorithms are the traditional STRUCTURE and CONCOR codes [37], which are used to study structural equivalence. REGE and its variations, focus on regular equivalence [36,38–40].

In the direct approach, the algorithm manipulates the elements of the network themselves, without needing an intermediate measure. The generalized blockmodeling, which is based on the direct approach, allows the user to pre-specify block types to which the algorithm should seek in the network. As the algorithm rearranges the nodes, a criterion function is optimized to reduce the number of inconsistencies (or error) that deviates from the pre-defined block [2]. The purposeful nature of generalized blockmodeling provides the researcher with a broad range of options that significantly enhance the chances of finding meaningful clusters in the network.

The traditional optimization of the criterion function in the generalized blockmodeling is computationally costly, and therefore it is not feasible for networks with hundreds of nodes [41]. This shortcoming may partially explain why academic papers with applications of generalized blockmodeling are scarce since real-world social network can quickly reach thousands of nodes. Studies have focused on networks of multiple relations [42,43]; valued networks [44], and divisions of labor in the job market [45]. Of specific interest to this work, is the blockmodeling of multi-mode networks. The nodes are classified into different categories, and ties between nodes of the same category are disregarded [32].

4. Results and discussions

4.1. The concept map

The base taxonomy maps thirty-four concepts in three major categories: construction activities, (boundary conditions and context) factors, and systems (typical project components) as shown in Fig. 2. Each concept is a node in the networks, and a tie is defined whenever there is an association between two nodes. Many relations are not logically possible or meaningful and therefore were removed from this study [5].

Reviewing recent works that studied the energy use in the construction phase [7,46,47], thirteen basic construction were identified. The second dimensions is the project components or physical systems, which are packages of items assembled or constructed during the project to form an asset of the project. They are divided into nine categories (Fig. 2), which were adapted from examples of O&G work breakdown structures (WBS) [48,49]. The third dimensions is the factors that impact the performance of construction activities. These factors serve as the boundary conditions of construction projects. Fig. 2 illustrates the three categories of factors that influence the energy use in construction activities. They encompass design factors, site characteristics, and resources. In principle, a factor may influence any construction activity and any other factor. Previous works [50,51] served as the basis for extracting relevant site factors (as shown in Fig. 2). For example, design parameters is a major factor that dictates the scope of construction activities as well as the type of resource to be used. This may include, [52–55] the selection of material or instruments; the nature of design code; and key design features (size, volume, thickness, flow, load, pressure, etc.). The resources used in construction is another important factor. They encompass the type, energy consumption levels of equipment; and their productivity and size [7,46,47].

4.2. The case projects

Project A is an industrial facility designed to manufacture DEF (diesel exhaust fluid), used in diesel engines to offset hazardous emissions. The owner has decided to build its DEF industrial plant in one of its existing fertilizer factories, which has produced urea and ammonia since the seventies. We interviewed eight participants. Their occupations at the time of the project were: planning engineers (2), project coordinator (1), commissioning coordinator (1), construction engineers (3) and contractor manager (1). The frequent topics in the responses reported by the participants were: the provision of a temporary shelter to reduce the weather impact; the mobilization of extra resources to make up for delays; the use of hydraulic jacks to erect the tanks and increase productivity; the adoption of pre-cast systems to enhance the productivity; the presence of existing underground structures (old foundations, drainage system, envelopes, etc.) that were not located in the basic design; and concerns about constructability of the basic design.

Projects B is an offshore regasification terminal that were part of a larger liquefied natural gas (LNG) project. Initially, the duration was considered feasible. However, weather conditions and rough seas during the winter hampered the project, leading to considerable delays. Seven participants were interviewed: a civil engineer, a quality assurance engineer, the field coordinator, two field managers, the commissioning coordinator, and the project manager. The frequent topics interviewees discussed: the impact of the rough sea conditions; the prefabrication of concrete parts and pipe spools as a strategy to reduce onsite work and energy; a poor seabed investigation that caused an impact on the deep foundation work; and a last-minute change in the design that reduced the number of piles of the structure of the pier, saving energy and time.

Project C is a 45-km onshore section of a natural gas pipeline that is also part of the flow system of the LNG project cited above. Seven participants were interviewed for this project: a field engineer, a design engineer, a planning engineer, two field managers, the field coordinator, and the commissioning coordinator. For this project, the most frequent topics were: the start of the construction activities in the rain season, which caused reworks, stoppages, low productivity and extra energy usage; the use of skidded equipment at stations as a modularized strategy to reduce the volume of onsite works; the risks involved in connecting the existing pipeline that was in operation and the new pipeline, which required the use of energy-intensive equipment for a longer time; and the lack of information about existing pipelines and other buried structures in the right-of-way, which led to delays, need for extra resources, and additional energy use.

4.3. A proposed generalized blockmodeling of three-mode networks

Networks can be viewed in a one-mode format, in which nodes have no categories. A two-mode network divides nodes into two categories and then build relationships between them. No ties between nodes of the same group is permitted or regarded relevant [56]. In our case, a three-mode network was the most suitable because the concept map has three categories. A classical matrix representation of three-mode directional networks is shown in Eq. (1) [57], where i, j and k are the number of nodes in each category (or mode). The index j represents the construction activities, which acts as the intermediate mode between the other two:

$$M = \begin{pmatrix} M_{ij} & M_{ik} \\ M_{jj} & M_{jk} \end{pmatrix} \tag{1}$$

The matrices M_{ij}, M_{ik} and M_{jk} represent the isolated incidence two-mode relations between each pair of modes. M_{jj} must be null since three-mode networks do not have ties between elements of the same mode. Besides, a restricted three-mode network, as in our case, is defined when there are no ties between nodes of the first and third modes [57]. Hence, a restricted network is presented as shown in Eq. (1). Similar to an incidence matrix, M is also a rectangular matrix with $i + j$ rows and $i + k$ columns. Although this convenient representation (known as Fararo's and Doreian's) may be useful for network analyses, restricted matrix is not suitable for blockmodeling three-mode networks since it artificially represents the nodes of the intermediate mode in the second row and in the first column simultaneously.

$$M = \begin{pmatrix} M_{ij} & 0 \\ 0 & M_{jk} \end{pmatrix} \tag{2}$$

The blockmodeling of two-mode networks locates the nodes of different modes in rows and columns separately. The process of blockmodeling of two-mode networks cannot be extended to three-mode level by simply using M as a two-mode proxy for a three-mode network. It is quite difficult to guarantee the blockmodeling algorithm will equally partition the nodes of the intermediate mode (construction activities) in the rows and columns. To guarantee the blockmodeling algorithm will not combine the nodes of any two modes into one cluster, we propose to use the augmented one-mode adjacency matrix [58], as shown in Eq. (3). Then, we use appropriate constraints so that the blockmodeling algorithm does not combine nodes of different categories.

$$\boldsymbol{Z} = \begin{pmatrix} 0 & \boldsymbol{D} & 0 \\ 0 & 0 & 0 \\ \boldsymbol{C} & 0 & 0 \end{pmatrix} \tag{3}$$

$\boldsymbol{C}$ and $\boldsymbol{D}$ are the incident matrices with the ties between factors and activities, and between activities and systems, respectively. The blocks in the diagonal are null because self-ties are not permitted in three-mode networks. The other remaining cells are also null because these relations were considered irrelevant or illogical; thus, they were removed from the scope of this research [5]. Eq. (4) explicitly states that the overall matrix has 34 × 34 cells; factors and activities sub matrix has 13 × 9; and activities and systems sub matrix has 12 × 13.

$$Z_{34\times34} = \begin{pmatrix} 0 & D_{13\times9} & 0 \\ 0 & 0 & 0 \\ C_{12\times13} & 0 & 0 \end{pmatrix} \tag{4}$$

The optimized blocks of Z is a three-mode blockmodeling problem based on a one-mode augmented matrix. To force discrete and clear selection of nodes from each mode, the user can specify that the nodes are permitted to stay only in clusters that were assigned to their corresponding categories by establishing a heavy penalty for such case (see

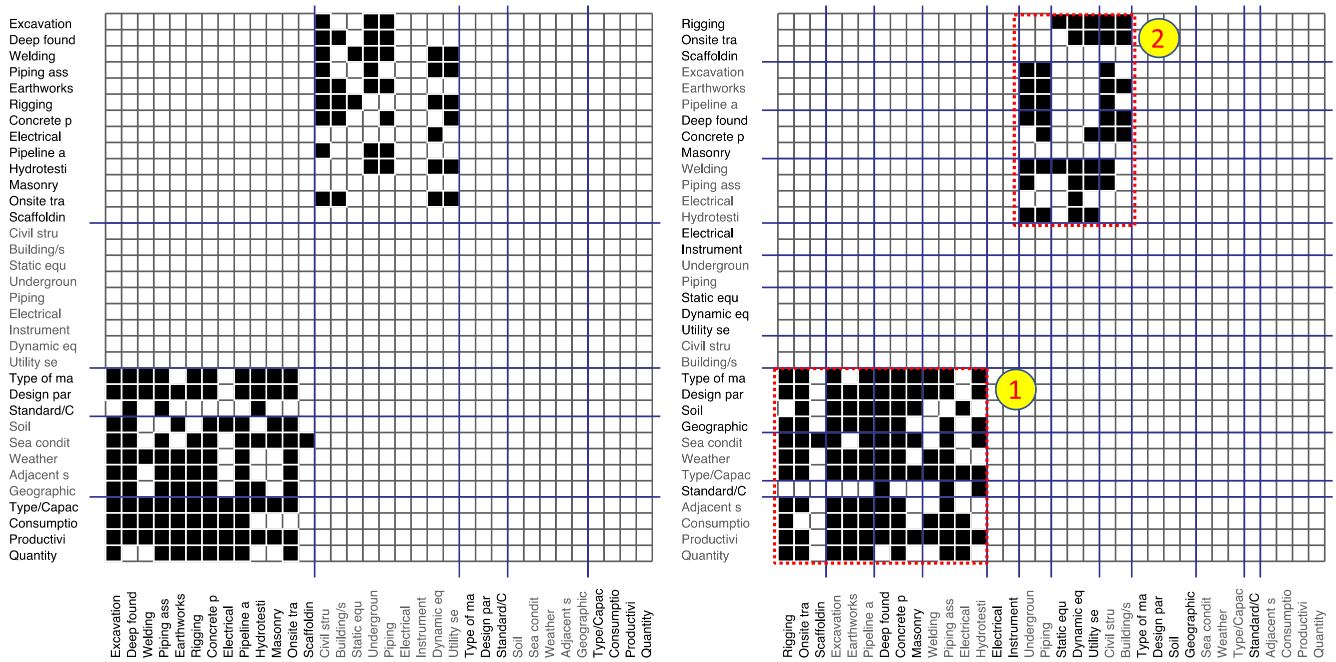

Fig. 3. The three-mode adjacency matrix of the benchmark network before (left) and after (right) a generalized blockmodeling with twelve clusters.

Appendix for how these constraints were implemented in Pajek [59], a non-commercial network analysis software).

Finally, the generic solution of the blockmodeling is presented in Eq. (5), in which the nodes representing factors, activities and systems are referred to by m, n and o, the number of clusters, respectively.

$$\boldsymbol{Z}^{*}=\begin{pmatrix} 0 & \cdots & 0 & D_{1,1} & \cdots & D_{1,n} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & D_{m,1} & \cdots & C_{m,n} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ C_{1,1} & \cdots & C_{1,m} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{o,1} & \cdots & C_{o,m} & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{5}$$

Defining the number of clusters (blocks) that the algorithm can detect in each mode is an iterative process. A set of initial exploratory runs are to be conducted to examine the range of feasible number of clusters. We started with 8 and 9 clusters in total. After several iterations, the number was increased to 12 because the first results did not provide a meaningful interpretation. Each category was grouped into clusters of 3 or 4 nodes.

In the case of generalized blockmodeling, the next step is to pre-specify the types of the blockmodels. Each block type defines a pattern of ties between the nodes. For example, in a complete block, all nodes are connected to each other. In a regular block, each node is connected to at least another node (for a detailed description of all block types, see [2]). The choice is not aleatory and should rely on previous knowledge of the network. In the *factor* × *activities* blocks, the factors are the source of the relations. Hence, one should look for at least row-regular blocks, in which all factors are connected to at least one activity. In the *activities* × *systems* blocks, each activity should be connected to at least one system, but also the systems should be connected to at least one activity. Hence, we set these blocks as regular. If a system is not connected to any activity, it should be put aside in a null block. Eq. (6) shows the permitted types of blocks in $\boldsymbol{Z}^{*}$.

$$\boldsymbol{Z}^{*}=\begin{pmatrix} 0 & \cdots & 0 & reg & \cdots & reg & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & reg & \cdots & reg & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ rre & \cdots & rre & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ rre & \cdots & rre & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{6}$$

Since the primary interest is to capture heterogeneous blocks (avoid the combination of nodes from the same category in the same cluster), we should establish the appropriate constraints to guarantee that the algorithm does not assign nodes of different categories in the same cluster. These constraints are also shown in the Appendix. We treat nodes from two different modes as exhaustively exclusive and orthogonal.

The block types defined in Eq. (6) are the least restrictive, for which a meaningful interpretation can be developed. It is important to allow the blockmodeling algorithm to also consider other (more restrictive or complete) types of blocks. In this case study, the allowable blocks across the modes were defined as follow:

- factors × activities: complete, null, regular and row-regular; they are abbreviated as *com*, *null*, *reg*, and *rre*, respectively.
- activities × systems: complete, null, and regular; they are abbreviated as *com*, *null*, and *reg*, respectively.

4.4. The benchmark network

To find a meaningful blockmodel for the generic network that can be compared with the networks of the case studies, four subgroups of activities were selected according to a functional classification: *logistics*, *heavy-duty equipment*, *structural civil works*, and *electro-mechanical assembly*. Onsite transportation, rigging, and scaffolding are the three activities that best represent the logistics in construction sites. Excavation, earthworks, and pipeline assembly are characterized by demanding pieces of heavy-duty equipment, such as excavators, dozers,

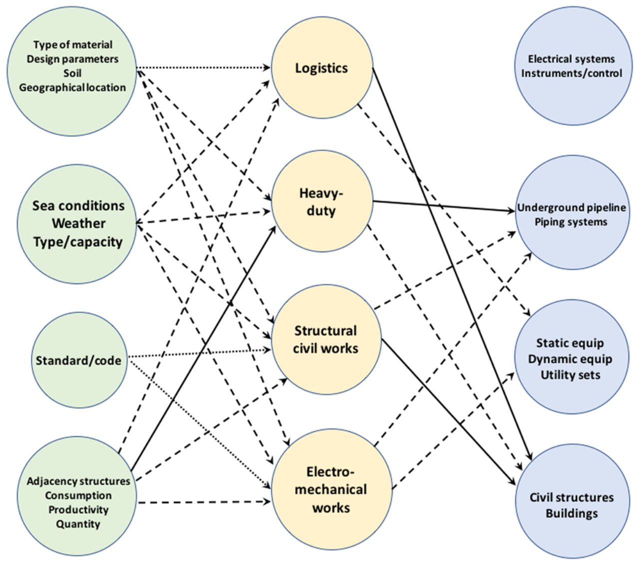

Fig. 4. Reduced graph of the generic network after the generalized blockmodeling.

graders, and pipelayers. Structural civil works encompass deep foundation, concrete preparation, and masonry works. Finally, electro-mechanical assembly, a significant activity in oil and gas projects, is comprised of electrical/instrument assembly, piping, welding, and hydrotesting. These four subcategories of activities form the base of clusters. The blockmodeling algorithm attempts to create blocks by permuting systems and factors to match the pre-specified clusters of activities according to the allowed blocks types and the constraints. Since the logistics, heavy-duty equipment, structural civil works, and electro-mechanical services have different characteristics, we expect that the final optimized blocks capture clusters of meaningful knowledge constructs that that better represent the generic network.

Fig. 3 shows the adjacency matrix of the benchmark network before and after the generalized blockmodeling. In the region of the factors vs. activities blocks (area “1”), there are four regular blocks. They are very close to complete blocks since there is only a missing tie in each one. They were classified by the algorithm as regular as this reduces the final error (a complete block with one error is a regular block without errors). This is one advantage of looking for several types of blocks in the generalized blockmodeling. In general, “factor × activities” blocks are quite dense. Fig. 4 shows the reduced graph of these blocks. Except for the sole standard/code block, every factor cluster is connected to all activity clusters, forming an almost perfect clique, even though the algorithm assigned less strict block types. A clique, in this case, suggests that every cluster of factors can influence the energy use in the logistics, heavy-equipment, structural works, and electro-mechanical assembly subgroups.

On the other hand, the interpretation of the “activities × systems” blocks (area “2” in Fig. 3) is significant different. First, electrical and instruments/control systems form a null block with every cluster of activities, which means that the participants did not recognize the construction of these systems as energy-intensive. The clusters “underground pipeline/piping systems” and “buildings/civil structures” have the highest degree (three), receiving contributions from sets of activities that reasonably match the disciplines involved in the construction of their systems. For example, buildings/civil structures form a complete block with structural civil works and a null block with electro-mechanical works. Interestingly, underground pipeline/piping systems forms a complete block with heavy-duty equipment activities with no influence from the “logistics” cluster, which may be seen as an inconsistency in the benchmark network as the construction of underground pipeline/piping is very logistics-intensive.

4.5. Comparing the case networks

In the following, the notation with apostrophe denotes the blocks, and the types of blocks of the benchmark network, whereas the absence of the apostrophe indicates case study networks. Each block type t in the case study networks can assume any of the four types described above. The block types in the benchmark network are denoted t'. The indexes o and p refers to the clusters in the rows and columns of the image matrix, respectively. The goal here is to compare each block of a case study network $t_{o,p}$with the corresponding block $t'_{o,p}$ in the benchmark network. To study to what extent a block in a case network is similar to a block in the same position in the benchmark network, Eq. (7) presents the block analogy index k.

$$k_{o,p} = \frac{r-e}{r}[(t_{o,p} = t'_{o,p}) \vee \{(t_{o,p} = com \vee t_{o,p} = reg) \wedge (t'_{o,p} = reg \vee t'_{o,p} = rre)\}] \tag{7}$$

where r is the number of possible ties, and e is the error (or the number of inconsistencies) observed in the block. The expression in brackets is an Iverson condition, so $k_{o,p}$ assumes any value between 0 and 1 if the expression is true, or 0 otherwise. The expression can be true in two cases: if the corresponding blocks in the case study and the generic network are the same ($t_{o,p} = t'_{o,p}$), or if the type of the observed block is more “complete” than the block in the generic network. For example, if $t_{o,p} = com$, and $t'_{o,p} = reg$, $k_{o,p} \neq 0$, since a regular block is a relaxation of a complete block. Generalizing to all blocks to be analyzed, the average block analogy index $\bar{k}$ is given by Eq. (8).

$$\bar{k} = \frac{1}{n}\sum_{\substack{o=1\\p=1}}^{n} k_{o,p} \tag{8}$$

In which n is the number of clusters to be analyzed. Note that $\bar{k}$ is a weighted average that accounts for (i) how identical the blocks of the benchmark and the case study network are, and (ii) what is the deviation (the error) from the ideal observed blocks. The average block analogy index $\bar{k}$ can assume any value between 0 and 1; therefore, the closer $\bar{k}$ is to 1, the more the blocks of the case study network matches the corresponding blocks in the generic network. Comparing the differences between blocks can pinpoint what is unique in a case network. The smaller $\bar{k}$, the more the case-based network moves away from the benchmark blockmodels. Table 1 presents the average block analogy index $\bar{k}$ for the three case studies. The first two columns are the $\bar{k}$ values for two groups of blocks: factors × activities and activities × systems. The last column is $\bar{k}$ for the entire network. Regarding the similarity of the blocks, Projects B and C present the largest $\bar{k}$, and the value of Project C is slightly lower. The values suggest that Project A is the one that least matches the blockmodels provided by the generic network. Projects B and C have similar $\bar{k}$, and, theoretically, they should provide less new/different knowledge than Project A.

(a) Project A

The interpretation of the distinct blocks is a manual process. Fig. 5 shows the adjacency matrix for the benchmark network and for Project A. The final image and error matrices for the evaluation can be found in Fig. 6. The highlighted blocks in red are showing which ones have

Table 1
Average block analogy index $\bar{k}$ for the “factor × activities” and “activities × systems” blocks.

Case study	$\bar{k}$ (factors × activities)	$\bar{k}$ (activities × systems)	$\bar{k}$ (overall)
Project A	0.328	0.438	0.383
Project B	0.401	0.549	0.475
Project C	0.401	0.556	0.478

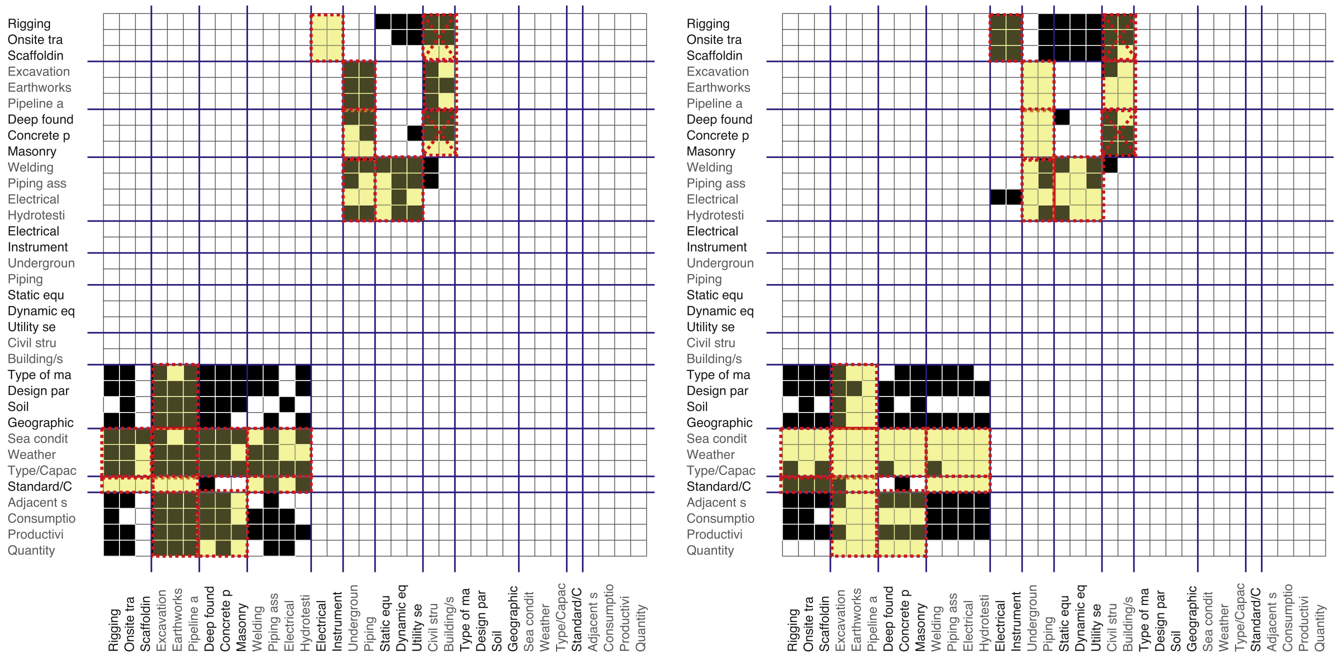

Fig. 5. The three-mode adjacency matrix of the blockmodel of the benchmark (left) and the rearranged Project A network (right).

different block types when compared with the benchmark network. The weather was a critical factor in Project A because of the temporary tents used to minimize the impact of the rainy season. This factor was purposefully neglected during the transformation from one-mode to three-mode. In the original one-mode network of Project A, the weather affects other factors, such as design parameters, productivity and type/capacity but no activity is influenced by the weather directly. Since ties between nodes of the same mode (categories) are not permitted in three-mode networks, this information was disregarded. This shortcoming, which is present in this study, should not occur if the data collection is designed for three-mode networks instead of one-mode. The factor "type/capacity" contributes to the only four inconsistencies of the four blocks. It significantly influenced the "logistics" cluster in this project because the contractor did not mobilize the resources with the right capacity to meet the demand of rigging and scaffolding. Two other energy-intensive activities, deep foundation and welding, were also affected by the contractors' decision on selecting resources type/capacity.

As opposed to a complete block in the benchmark network, the cluster "type of material/equipment, design parameters, soil, geographical location" only forms a row-regular block with the cluster "excavation, earthworks, pipeline assembly". This reflects the fact that there was no significant scope of earthworks and pipeline works in Project A. It is a row-regular block because it forms a complete column of relations with excavation, a very energy-intensive and challenging activity in Project A. The presence of pre-existing structures located underground, which had to be removed, impacted the cost and energy.

In the benchmark matrix, a standard or a code influence on energy use was low. In project A, the owner's internal standards highly influenced the scope of the project. This is also observed in the complete block formed with the "logistics" cluster. Despite being a regular block in the benchmark network, the cluster "adjacency structures, consumption, productivity, quantity" was classified as null with the cluster "structural civil works" and a large number of inconsistencies. Curiously, these inconsistencies happen with two important factors in Project A: adjacency structures and productivity. The former was the cause of so many impacts during many activities as discussed above, and the latter is associated with the acceleration plan that mobilized extra resources but sacrificed the productivity, as well as and energy efficiency.

Regarding the sets of blocks in the "*activities* × *systems*" sub matrix, the first divergence is the complete block formed between the "logistics" activities and the "electrical/instruments" systems. The logistics significantly affected most of the systems, even during the construction of electrical/instruments/control systems. Project A took place at an existing site in operation with an old infrastructure that impacted the integration between the new and the existing facilities, which caused an extra energy expenditure. Given that Project A did not have a large scope involving earthmoving and pipeline assembly, the cluster with the "heavy-equipment" activities just form null blocks with all the four clusters of systems. Hence, the contribution of this cluster on the energy use was not perceived as important in this project. There is one

	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	com	-	com	reg	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	reg	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	reg	rre	reg	reg	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-
11	com	rre	rre		-	-	-	-	-	-	-	-
12	reg	-	-	com	-	-	-	-	-	-	-	-

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	3	0	0	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	2	3	4	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	4	0	0	0	0	0	0	0	0
10	2	0	1	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	3	2	6	4	0	0	0	0	0	0	0	0

Fig. 6. Image and error matrices for the evaluation of Project A.

inconsistency, however, concerning excavation and civil structures because of the extra mobilized excavators and hydraulic hammers to remove the old concrete found underground.

The cluster related to structural civil works just forms a non-null block with the cluster "civil structures/buildings" systems. Although a null block, there is an important inconsistency with the cluster "static equipment, dynamic equipment, utility sets". That is a tie between deep foundation and static equipment caused by the problems in the foundation of the storage tanks, which were in the critical path of the construction schedule. The contractor had to mobilize extra resources to remove and store the contaminated mud temporarily. Finally, there are two different blocks formed by the cluster of electro-mechanical works: the first is with the cluster "underground pipeline/piping", and the second is with the cluster "static/dynamic/utility sets". Both blocks have few inconsistencies, and all are associated to systems that had significant influence on the performance of the project, such as a high volume of stainless steel piping, and a large scope of utility sets and static equipment (the storage tanks).

(b) Project B

Fig. 7 shows the blockmodel in the form of the adjacency matrix for both the benchmark network and Project B (the final image and error matrices of Project B can be found in Fig. 10, Appendix). The results suggest Projects A and B have several common features because many of their blocks diverge equally from the benchmark network. They have similarities even though the former is an industrial facility and the latter is an offshore LNG terminal. Project B is an offshore terminal whose construction was significantly affected by the sea conditions, the location of the offsite facilities, and the challenging logistics. As with Project A, the "logistics" cluster also forms a block with the "electrical/instrument" system, and this relation does not appear in the blockmodel of the benchmark network. The blocks formed by the four clusters of activities and the cluster "underground pipeline, piping" are also null with some inconsistencies that highlight the importance of piping systems in the project due to a large scope of stainless steel pipes. A null block is also present between the "heavy-duty" activities and the "civil structures/buildings" clusters with an important inconsistency that points out the challenges of the excavation during the construction of the structure of the terminal due to poor soil investigation that caused reworks and extra energy spending.

In the set of blocks formed by factors and activities, some blocks are row-regular or null with inconsistencies, as opposed to the more regular and complete blocks in the benchmark network. All four clusters of factors form null blocks with the "heavy-duty" activities because two of these activities are not present in offshore projects: earthworks and pipeline assembly. However, as mentioned above, the excavation, a critical activity in Project B, produced a few inconsistencies in the blocks. Regarding the clusters of "logistics" and the "type of material/equipment, design parameters, soil, geographical location", some changes in the design of the pier allowed the contractor to pre-cast some parts of the structure in the onshore site. This decision had profound implications on the logistics of the project, especially rigging and onsite transportation.

(c) Project C

Project C was the construction of a natural gas pipeline. This kind of project is characterized by a significant influence of the logistics on the performance as well as the mobilization of energy-intensive equipment such as excavators, bulldozers, and sideboom pipelayers. As with the other cases, Fig. 8 presents the adjacency matrix, which was permutated to fit the blockmodels of the benchmark network (the final image and error matrices of Project C can be found in Fig. 11, Appendix). We can notice in the figure the importance of the "logistics" activities because it forms a complete block with all the cluster systems but one: static/dynamic equipment, utility sets. The weight of these systems in the scope of pipeline projects is not significant. However, the most relevant here are the inconsistencies of the blocks. First, scaffolding has no connection because pipeline services hardly have work at heights. Second, Rigging and onsite transportation have two ties with "utility sets" system because the project team had decided to invest in modularized utility equipment, which simplified the logistics in the field.

The modularization of utility sets was reflected in the blocks. The cluster of electro-mechanical activities is not an ideal null block with the cluster "static/dynamic equipment/utility sets" because of the existing relations with the "utility set" node. Investing in modularized equipment significantly reduced the volume of services regarding welding, piping and electrical assembly and therefore less energy was consumed in the construction.

Regarding the sets of nodes in the blocks "*factor* × *activities*", the inconsistencies in the null blocks formed by the cluster "weather, sea

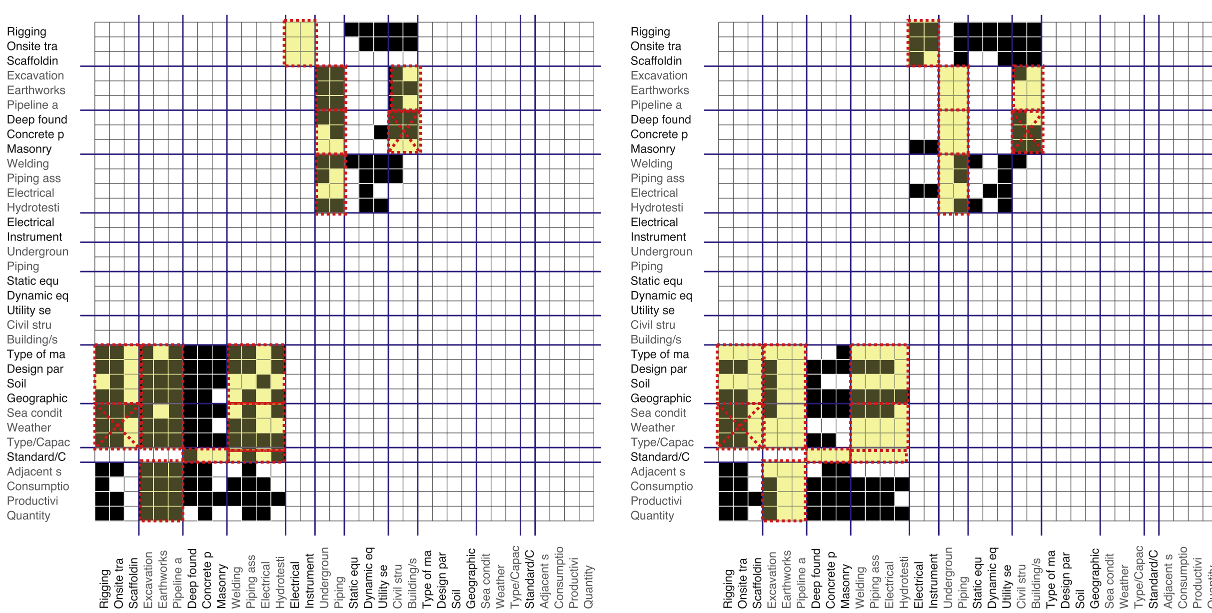

Fig. 7. The three-mode adjacency matrix of the blockmodel of the benchmark (left) and the rearranged Project B network (right).

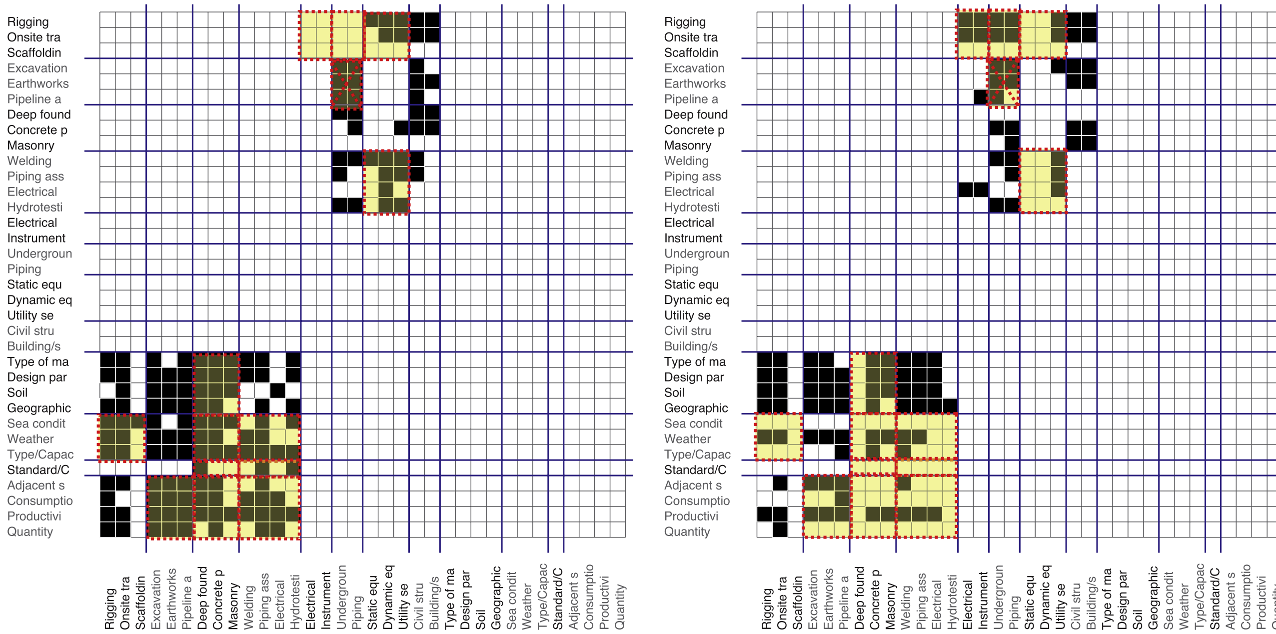

Fig. 8. The three-mode adjacency matrix of the blockmodel of the benchmark network (left) and the arranged Project C network (right).

conditions, type capacity" provides some interesting insights. The decision of the contractor to postpone part of the schedule for the rainy season had several negative consequences, which are represented in the form of the ties with many important activities. Also, the row-regular block containing "adjacency structures" and "productivity" accounts for the impact caused by the inadequate investigation of the right-of-way that could have detected the details of existing pipelines in the region. This unexpected problem caused delays and low productivity. For the blocks formed by the factors and the clusters "structural civil works" and "electro-mechanical assembly", two observations can be made. The first is the absence of ties with deep foundation because this activity was not part of the scope. The second is the presence of ties involving the factors "weather, productivity" and "concrete preparation, masonry, welding, piping and electrical assembly". The concrete preparation is an important process in pipeline constructions, and the location of the temporary facilities highlighted its significance. Welding and piping assembly are quite important in the construction of the aboveground facilities. The productivity of two main aboveground facilities was significantly affected by the weather, and these relations are important inconsistencies in the null blocks of the electro-mechanical cluster of activities.

4.6. The adjusted average block analogy index $\bar{k}$

We classified the blocks and calculated the error by performing a "dummy" blockmodeling with the partition of the ideal blockmodels from the benchmark network. Table 2 shows the adjusted values of $\bar{k}$ calculated by using Eqs. (7) and (8). The blockmodeling algorithm attempts to minimize the errors. For example, in Project A, the blocks formed by the "logistics" and the "structural civil works" clusters, and

Table 2
Adjusted average block analogy index $\bar{k}$ for the "factor × activities" and "activities × systems" blocks.

Case study	$\bar{k}$ (factors × activities)	$\bar{k}$ (activities × systems)	$\bar{k}$ (overall)
Project A	0.328	0.543	0.435
Project B	0.443	0.602	0.522
Project C	0.401	0.608	0.504

the cluster "civil structures/buildings" systems were classified as perfect regular, as opposed to complete blocks in the ideal blockmodel of the benchmark network. However, these blocks could have been classified as complete with only one error each as well. If the algorithm had classified these blocks this way, $\bar{k}$ would be slightly higher than the value provided in Table 1. This shortcoming suggests that a correction is necessary for $\bar{k}$ to best represent the analogy between the blocks. As such, during the manual interpretation of the blocks of the case-based network, we detected the blocks that were considered divergent because of this technicality and changed their classification manually to the type that best fits the block. These blocks are identified with a red cross in Figs. 5, 7 and 8. The adjusted image and error matrices are shown in Fig. 9 for all the case studies.

5. Conclusion

Blockmodeling was used to extract concept clusters from semantic networks, which was generated based on modeling the characteristics/knowledge gained in a project. Using concept blocks to capture and compare concept networks is more practical than a node-by-node analysis. Blockmodels provide a more meaningful way of grouping the nodes and finding their interrelations. $\bar{k}$, an index of deviations between a set of blocks, was used to compare the concept networks of three case studies. Knowledge analysts/managers can study to what extent a project network is similar to the benchmark network. They can visualize, through the adjacency matrices, the differences in their structures as well. Besides, each cluster has a meaning with regards to a group of activities, associated factors, and affected common systems.

Generalized blockmodeling has a significant number of parameters and constraints that influence the number of solutions. Although the number of inconsistencies in each solution is not the most crucial aspect, each of them has to be manually interpreted before a decision is made about which solution (size and nature of the blocks) is more meaningful. Depending on the number of possibilities and solutions, this trial-and-error process may be time-consuming. Therefore, the seed blocks must be carefully selected.

As the routine of blockmodeling in this research includes only a conventional optimization method, the authors encourage future works that attempt to implement three-mode generalized blockmodeling with evolutionary optimization tools [41]. This process should guarantee

	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	com	-	com	com	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	com	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	reg	rre	reg	reg	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-
11	com	rre	rre	-	-	-	-	-	-	-	-	-
12	reg	-	-	com	-	-	-	-	-	-	-	-

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	3	0	1	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	2	3	4	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	4	0	0	0	0	0	0	0	0
10	2	0	1	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	3	2	6	4	0	0	0	0	0	0	0	0

	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	reg	-	reg	com	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	com	-	-	-	-
4	-	-	-	-	-	-	reg	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	reg	-	-	-	-	-	-	-	-	-
10	reg	-	reg	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-
12	reg	-	reg	reg	-	-	-	-	-	-	-	-

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	3	0	0	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	2	0	0	1	0	0	0	0
4	0	0	0	0	2	3	0	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	4	3	0	7	0	0	0	0	0	0	0	0
10	3	1	3	3	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	0	3	0	4	0	0	0	0	0	0	0	0

	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	com	com	-	com	-	-	-	-
2	-	-	-	-	-	com	-	com	-	-	-	-
3	-	-	-	-	-	reg	-	com	-	-	-	-
4	-	-	-	-	-	reg	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-
9	rre	reg	rre	reg	-	-	-	-	-	-	-	-
10	-	reg	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-
12	rre	reg	-	-	-	-	-	-	-	-	-	-

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	2	2	2	2	0	0	0	0
2	0	0	0	0	1	1	1	2	0	0	0	0
3	0	0	0	0	0	2	0	2	0	0	0	0
4	0	0	0	0	2	2	3	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	2	3	3	3	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	3	3	2	4	0	0	0	0	0	0	0	0

Fig. 9. Adjusted image (left) and error (right) matrices for the evaluation of Project A, B and C (from top to bottom) based on the ideal blockmodels of the benchmark network.

that only global minimum solutions are found, reduce the time of each run, and perhaps reduce the number of solutions. During the proposed analysis, the authors focused on comparing the inconsistencies between blocks in different networks. In future research, an alternative (and perhaps complementary) approach could be beneficial: what are the inconsistencies that have prevented blocks of the same type from being ideally equal. Besides, future works can attempt to implement automated knowledge retrieval and representation tools to generate the case text and the concept networks.

Acknowledgements

The authors of this paper are deeply grateful to support from the University of Toronto and Petrobras – Petroleo Brasileiro S.A – for supporting this research.

Appendix A

Defining the constraints for the blockmodeling of the benchmark network in Pajek

Using an augmented adjacency matrix to blockmodel a three-mode network requires defining some constraints to avoid clusters of mixing nodes of different modes or categories. There are two ways to define a blockmodel in Pajek: entering the data directly into the blockmodeling toolbox or writing an mdl file. An mdl file, a text file with a mdl extension, is employed whenever the user needs to run the same blockmodel several times. We show how the constraints were defined using the syntax required in mdl files in Pajek. For more details on how to implement a complete mdl file, see the reference manual of Pajek. There are several types of constraints in Pajek, but we only show the three types used in this study. The general syntax for a constraint is given by:

c penalty y z

with the following forms:

1. If $c = 1$: $y \in C_z$ – vertex y belongs to cluster C_z;
2. If $c = 2$: $y \in C_z$ – vertex y does not belong to cluster C_z;
3. If $c = 6$: $y \geq |C_z|$ – cluster C_z contains at most y vertices.

To pre-define the four clusters of activities, we used the first form for nodes 1–13 (please refer to Fig. 2 for the ID of each node). To inform Pajek that it must avoid grouping factors and systems in the same cluster, we used the second form for the nodes representing the systems (14–22) and the factors (23–34). To guarantee the clusters do not contain more than four nodes, the third form is the most suitable. In each form, penalties for violations are placed to support the discovery of more consistent blocks. We used a value of 100 for first and second forms, and 50 for the third. The reason for a lower penalty for the third form was that to avoid being strict about the size of the clusters.

*CONSTRAINTS
[Pre-defining the clusters of activities]
1 100 6 1
1 100 12 1
1 100 13 1
1 100 1 2
1 100 5 2
1 100 9 2
1 100 2 3
1 100 7 3
1 100 11 3
1 100 3 4
1 100 8 4
1 100 4 4
1 100 10 4
[Constraints to avoid grouping systems with factors]
2 100 14 9
2 100 15 9
.
.
.
2 100 22 9
2 100 14 10
2 100 15 10
.
.
.

[cont.]
2 100 22 10
2 100 14 11
2 100 15 11
.
.
.
2 100 22 11
2 100 14 12
2 100 15 12
.
.
.
2 100 22 12
[Constraints to avoid grouping factors with systems]
2 100 23 5
2 100 24 5
.
.
.
2 100 34 5
2 100 23 6
2 100 24 6
.
.
.
2 100 34 6

2 100 23 7

[cont.]
2 100 24 7
.
.
.
2 100 34 7
2 100 23 8
2 100 24 8
.
.
.
2 100 34 8
[Constraints to define the maximum number of nodes in each cluster]
6 50 3 1
6 50 3 2
6 50 3 3
6 50 4 4
6 50 4 5
6 50 4 6
6 50 4 7
6 50 4 8
6 50 4 9
6 50 4 10
6 50 4 11
6 50 4 12
*EOM

The image and error matrices of projects B and C are shown below:

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aei.2019.01.003.

References

[1] J. Boyd-Graber, D. Mimno, D. Newman, Care and feeding of topic models: problems, diagnostics, and improvements, Handb. Mix. Membsh. Model. Their Appl. 225255 (2014).
[2] P. Doreian, V. Batagelj, A. Ferligoj, Generalized Blockmodeling, first ed., Cambridge University Press, New York City, NY, 2005.
[3] P. Doreian, V. Batagelj, A. Ferligoj, Generalized blockmodeling of two-mode network data, Soc. Netw. 26 (2004) 29–53, https://doi.org/10.1016/j.socnet.2004.01.002.
[4] T.E. El-Diraby, Domain ontology for construction knowledge, J. Constr. Eng. Manag. 139 (2013) 768–784.
[5] R.R. Aragao, T.E. El-Diraby, Using network analytics to capture knowledge: three cases in collaborative energy-oriented planning for oil and gas facilities, J. Clean. Prod. 209 (2019) 1429–1444, https://doi.org/10.1016/j.jclepro.2018.10.346.
[6] 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.
[7] S. Junnila, A. Horvath, A.A. Guggemos, Life-cycle assessment of office buildings in Europe and the United States, J. Infrastruct. Syst. 12 (2006) 10–17, https://doi.org/10.1061/(ASCE)1076-0342(2006) 12:1(10).
[8] T. Hong, C. Ji, M. Jang, H. Park, Assessment model for energy consumption and greenhouse gas emissions during building construction, J. Manag. Eng. 30 (2014) 226–235, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000199.
[9] P.J. Davies, S. Emmitt, S.K. Firth, Challenges for capturing and assessing initial embodied energy: a contractor's perspective, Constr. Manag. Econ. 32 (2014) 290–308, https://doi.org/10.1080/01446193.2014.884280.
[10] G.J. Treloar, R. Fay, B. Ilozor, P.E.D. Love, An analysis of the embodied energy of office buildings by height, Facilities 19 (2001) 204–214, https://doi.org/10.1108/02632770110387797.
[11] Construction Industry Institute (CII), Effective Management Practices and Technologies for Lessons Learned Programs, 2007.
[12] 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.
[13] A. Singhal, Introducing the Knowledge Graph: things, not strings, Google, 2012. < https://googleblog.blogspot.ca/2012/05/introducing-knowledge-graph-things-not.html > (accessed October 2, 2017).
[14] 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. doi: 10.1234/12345678.
[15] 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. doi: 10.1145/988672.988687.
[16] J. Read, Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification, in: ACL Student Res. Work., Association for Computational Linguistics, 2005, pp. 43–48. doi: 10.3115/1628960.1628969.
[17] M. Park, K.W. Lee, H.S. Lee, P. Jiayi, J. Yu, Ontology-based construction knowledge retrieval system, KSCE J. Civ. Eng. 17 (2013) 1654–1663, https://doi.org/10.1007/s12205-013-1155-6.
[18] N.M. El-Gohary, T.E. El-Diraby, Domain ontology for processes in infrastructure and construction, J. Constr. Eng. Manag. 136 (2010) 730–744, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000178.
[19] J.F. Sowa, Semantic networks, in: Encycl. Artif. Intell., second ed., Wiley-Interscience, 1992. doi: 10.1177/089431849901200420.
[20] B. Harrington, ASKNet: Automatically Creating Semantic Knowledge Networks from Natural Language Text, University of Oxford, 2009.
[21] P.R. Wojtinnek, B. Harrington, S. Rudolph, S. Pulman, Conceptual knowledge acquisition using automatically generated large-scale semantic networks, in: 18th Int. Conf. Concept. Struct. ICCS 2010, Springer-Verlag, Kuching, Sarawak, Malaysia, 2010, pp. 203–206. doi: 10.1007/978-3-642-14197-3_23.
[22] A. Freitas, D.S. Carvalho, J.C.P. da Silva, S. O'Riain, E. Curry, A semantic best-effort approach for extracting structured discourse graphs from Wikipedia, in: Proceeding 1st Work. Web Linked Entities, (ISWC) 1st Work. Web Linked Entities, (ISWC), 2012, p. 12.
[23] M. Oliveira, J. Gama, An overview of social network analysis, Wiley Interdiscip Rev. Data Min. Knowl. Discov. 2 (2012) 99–115, https://doi.org/10.1002/widm.1048.
[24] P. Chinowsky, J. Diekmann, V. Galotti, Social network model of construction, J. Constr. Eng. Manag. 134 (2008) 804–812, https://doi.org/10.1061/(ASCE)0733-9364(2008) 134:10(804).
[25] M.N. Bakht, Analyzing Social Network Discussions of Infrastructure Projects – Towards a Bottom-Up Decision Making Environment, University of Toronto, 2015.
[26] M.N. Bakht, T.E. El-Diraby, Hidden social networks that drive online public involvement in infrastructure construction case study of light rail transit projects in North America, Transp. Res. Rec. (2014) 84–93, https://doi.org/10.3141/2453-11.
[27] A.P.C. Chan, A.A. Javed, F.K.W. Wong, C.K.H. Hon, H. Zahoor, S. Lyu, The application of social network analysis in the construction industry of Hong Kong, in: 1st Int. Conf. Emerg. Trends Eng. Manag. Sci., Queensland University of Technology, Peshawar, Pakistan, 2014, p. 31. < http://www.cusit.edu.pk/icetems/Proceeding/Civil Engineering/HK-CE-10.pdf%5Cnhttp://eprints.qut.edu.au/80109/3/80109.pdf > .

[28] R. Alsamadani, M. Hallowell, A.N. Javernick-will, Measuring and modelling safety communication in small work crews in the US using social network analysis, Constr. Manag. Econ. 31 (2013) 568–579, https://doi.org/10.1080/01446193.2011.685486.
[29] H. Lingard, P. Pirzadeh, N. Blismas, R. Wakefield, B. Kleiner, Exploring the link between early constructor involvement in project decision-making and the efficacy of health and safety risk control, Constr. Manag. Econ. 32 (2014) 918–931, https://doi.org/10.1080/01446193.2014.911931.
[30] A. Woldesenbet, H.D. Jeong, H. Park, Framework for integrating and assessing highway infrastructure data, J. Manag. Eng. 32 (2015) 4015028, https://doi.org/10.1061/(ASCE)ME.1943-5479.0000389.
[31] W. de Nooy, A. Mrvar, V. Batagelj, Exploratory Social Network Analysis with Pajek, first ed., Cambridge University Press, New York, 2005, , https://doi.org/10.1017/CBO9781107415324.004.
[32] S.P. Borgatti, M.G. Everett, Regular blockmodels of multiway, multimode matrices, Soc. Netw. 14 (1992) 91–120, https://doi.org/10.1016/0378-8733(92)90015-Y.
[33] S. Wasserman, K. Faust, Social Network Analysis: Methods and Application, first ed., Cambridge University Press, New York, 1994.
[34] R.L. Breiger, S.A. Boorman, P. Arabie, An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling, J. Math. Psychol. 12 (1975) 328–383, https://doi.org/10.1016/0022-2496(75)90028-0.
[35] R.S. Burt, Positions in Networks, Soc. Forces. 55 (1976) 93–122.
[36] S.P. Borgatti, M.G. Everett, Two algorithms for computing regular equivalence, Soc. Netw. 15 (1993) 361–376, https://doi.org/10.1016/0378-8733(93)90012-A.
[37] V. Batagelj, A. Ferligoj, P. Doreian, Direct and indirect methods for structural equivalence, Soc. Netw. 14 (1992) 63–90, https://doi.org/10.1016/0378-8733(92)90014-X.
[38] M.G. Everett, S.P. Borgatti, Regular equivalence: general theory, J. Math. Sociol. 19 (1994) 29–52.
[39] A. Ziberna, Direct and indirect approaches to blockmodeling of valued networks in terms of regular equivalence, J. Math. Sociol. 32 (2008) 57–84, https://doi.org/10.1080/00222500701790207.
[40] A. Žiberna, Evaluation of Direct and Indirect Blockmodeling of Regular Equivalence in Valued Networks by Simulations, 2009.
[41] 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.
[42] M. Dabkowski, R. Breiger, F. Szidarovszky, Simultaneous-direct blockmodeling for multiple relations in Pajek, Soc. Netw. 40 (2015) 1–16, https://doi.org/10.1016/j.socnet.2014.06.003.
[43] A. Žiberna, Blockmodeling of multilevel networks, Soc. Netw. 39 (2014) 46–61, https://doi.org/10.1016/j.socnet.2014.04.002.
[44] A. Žiberna, Generalized blockmodeling of valued networks, Soc. Netw. 29 (2007) 105–126, https://doi.org/10.1016/j.socnet.2006.04.002.
[45] J. Glückler, R. Panitz, Unpacking social divisions of labor in markets: Generalized blockmodeling and the network boom in stock photography, Soc. Netw. 47 (2016) 156–166, https://doi.org/10.1016/j.socnet.2016.07.002.
[46] M.M. Bilec, A Hybrid Life Cycle Assessment Model for Construction Processes, University of Pittsburgh, 2007.
[47] A.A. Guggemos, A. Horvath, Comparison of environmental effects of steel- and concrete-framed buildings, J. Infrastruct. Syst. 11 (2005) 93–101, https://doi.org/10.1061/(ASCE)1076-0342(2005) 11:2(93).
[48] M.A. El-Reedy, Project Management in the Oil and Gas Industry, Scrivener Publishing LLC, Salem, Massachusetts, 2016.
[49] A.B. Badiru, S.O. Osisanya, Project Management for the Oild and Gas Industry – A World System Approach, first ed., Taylor & Francis Group, Boca Raton, FL, 2013.
[50] G. Ding, P.J. Forsythe, Sustainable construction: life cycle energy analysis of construction on sloping sites for residential buildings, Constr. Manag. Econ. 31 (2013) 254–265, https://doi.org/10.1080/01446193.2012.761716.
[51] P.O. Akanni, A.E. Oke, O.A. Akpomiemie, Impact of environmental factors on building project performance in Delta State, Nigeria, HBRC J. 11 (2015) 91–97, https://doi.org/10.1016/j.hbrcj.2014.02.010.
[52] S.A. Assaf, M. Al-Khalil, M. Al-Hazmi, Causes of delay in large building construction projects, J. Manag. Eng. 11 (1995) 45–50, https://doi.org/10.1061/(ASCE)0742-597X(1995) 11:2(45).
[53] S.A. Assaf, S. Al-Hejji, Causes of delay in large construction projects, Int. J. Proj. Manag. 24 (2006) 349–357, https://doi.org/10.1016/j.ijproman.2005.11.010.
[54] M.H. Fallahnejad, Delay causes in Iran gas pipeline projects, Int. J. Proj. Manag. 31 (2013) 136–146, https://doi.org/10.1016/j.ijproman.2012.06.003.
[55] Construction Industry Institute (CII), Constructability – A Primer, University of Texas at Austin, 1986.
[56] S.P. Borgatti, M.G. Everett, Network analysis of 2-mode data, Soc. Netw. 19 (1997) 243–269.
[57] T.J. Fararo, P. Doreian, Tripartite structural analysis: generalizing the Breiger-Wilson formalism, Soc. Netw. 6 (1984) 141–175.
[58] D. Melamed, R.L. Breiger, A.J. West, Community structure in multi-mode networks: applying an eigenspectrum approach, Connections 33 (2013) 18–23 http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=90250256&site=ehost-live.
[59] A. Mrvar, V. Batagelj, Pajek, Pajek XXL, Pajek 3XL – Programs for Analysis and Visualization of Very Large Networks – Reference Manual, Ljubljana, 2003. < vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf > .