An Approach to Data Driven Process Discovery in the Cost Estimation Process of a Construction Company

Tobias Kropp[1], Alexander Bombeck[1] and Kunibert Lennerts[1]

[1]Institute of Technology and Management in Construction, Karlsruhe Institute of Technology (KIT), Germany

tobias.kropp@kit.edu, alexander.bombeck@kit.edu, kunibert.lennerts@kit.edu

Abstract -

This work examines the potential of process mining in the Architecture, Engineering, Construction, and Operation (AECO) industry, where process mining is rarely applied. The main reason is that standardised processes are hardly ever performed due to the complexity of projects. To address this application gap, the software-supported cost estimation process for a tender in a German construction company is examined. For this purpose, data sets from three different projects are exported from the software *RIB iTWO*^{TM} and analysed with regard to process discovery. Investigations are carried out from the control flow, case and organisational perspectives. A particular problem in the analysing part is the currently inadequate quality of the data sets to be examined. Due to data quality issues the consideration of temporal aspects is rather not possible. Therefore, consistent and appropriate forms of logging must be implemented in software systems that are used to support the AECO industry. For the present use case an automated working script was created to process the log data to fix further quality issues and meet process mining requirements.

The results show that it is essential to establish standardised language rules that are generally valid throughout the industry and that are subsequently referenced by all utilised software systems. This enables comparative analyses across projects and companies to make process mining methods become routinely and profitably applicable in the AECO industry.

Keywords -

Process Mining; Process Discovery; AECO Industry; Civil Engineering; Building Construction; Cost Estimation

1 Introduction

Process mining is to be seen as a link between data science and process science since it brings together traditional model-based process analysis and data-centric analysis techniques [1]. While several process mining applications can be mostly found in the areas of financials, healthcare or manufacturing, the method is rarely applied in the Architecture, Engineering, Construction, and Operation (AECO) industry [2]. In the planning phase within the AECO industry, there are usually complex and individual projects handled [3]. Standardised processes are rarely found. Tasks that require creativity and problem-solving skills may lead to unstructured process flows and usually challenge the applicability of process mining methods [1].

Nevertheless, process-supporting software is also increasingly being used in the AECO industry. Countless data is already collected in everyday business, but is hardly ever analysed. It is necessary to verify that meaningful interpretations of the collected data can be implemented with the help of process mining. Currently, only a few research is conducted on the use of process mining methods in the AECO industry. This emphasises the importance of the scientific work highlighted in this paper, that addresses the cost estimation process of a construction company.

In this paper, Section 2 reviews available scientific studies on process mining applications in the AECO industry and underlines the sector specific application gap. Section 3 describes the fundamentals of process mining and the case study from this paper. In Section 4 the investigated use case is presented. Section 5 shows the critical conclusions and recommends topics to be addressed in future work.

2 Related Work

Overall, only a few research studies are done on the use of process mining methods in connection with the AECO industry.

In [4] two disciplines of process mining, process discovery and conformance checking, are used to analyse process steps in modular construction through RFID tracking data. The case study is conducted as a laboratory experiment and can generally be assigned to a rather structured manufacturing process than to the classical AECO industry.

In the context of [5] three case studies in the field of civil engineering are investigated to cover different phases of civil engineering projects. The first case study is about the planning phase of a civil road construction project with data from a project management tool to find loops and bottlenecks in the execution. In the second case study data from drone images documenting construction progress in the form of point cloud models (as-built models) are compared with planned (as-planned) Building Information Modeling (BIM) models. The third case study was in the context of facility management using maintenance (error) data.

Some more scientific work can be found in the context of BIM with [6, 7, 8, 9, 10]. All of them treat real project or laboratory data from the design authoring tool *Autodesk*

Revit™ [11]. The in the present work investigated software system, *RIB iTWO™*, can also be used in the context of BIM, for example to take over design authoring tool data to perform automatically mass determination of building elements.

Zhang et al. [6] aim to provide a productivity measurement in the workflow of different BIM design authors. Therefore, it concentrates on the most frequent patterns (sequence of commands) within all the log data from real life projects to assess the process duration of those patterns. The work of Yarmohammadi et al. [7] addresses the same topic but just examines patterns within real project data with the most frequent activities without considering time aspects. Other scientific work in the context of BIM like [5, 9, 10] focus rather on a complete (from starting point to ending point) process view over process instances with respect to time aspects. The latter is more common as an holistic process mining approach.

Two of the authors from [6] conducted another case study in [8] on real data to discover social networks in BIM-based collaborative design and furthermore confirm findings through interviews with several design project managers of the related company.

Kouhestani and Nik-Bakht [9] aim to use log data which have arisen from staged processes to measure the performance of the project teams with focusing on the design authors as cases to track their individual workflow. It is already outlined that the AECO industry has in general a lack of log data collections due to the fact that digital planning methods are just recently about to be applied with BIM methods [9]. The same authors in a later work [10] present firstly the same case study as [9] but utilise building elements as cases instead of the design authors to discover and analyse the element centered processes completed by the design team during design authoring. A second case study considers a bigger data set from again staged processes, and process conformance analysis is furthermore carried out in addition to process discovery.

In [12] a case study on Engineering Change Request (ECR) is examined to explain the importance of preprocessing the event logs before importing them into process mining software. The investigated data set covered only 58 cases and thus represents a small data base. Nevertheless, it is already shown too, that there can be quality problems in the context of log data in the AECO industry, which require extensive processing of the data.

In summary, there is only a small number of case studies that deal with process mining in the AECO industry, most of which focus on design authoring in the context of BIM. To address the often criticised data quality issues and make the application of process mining methods in the AECO industry more feasible, the present work examines the software-supported cost estimation process in a German construction company. This paper follows the similar objectives as [12], as it mainly addresses data pre-processing obstacles, but furthermore conducts initial analysis and interpretation steps.

3 Fundamentals of Process Mining and Adopted Approach

In the Process Mining Manifesto [13], written by members and supporters of the IEEE Task Force on Process Mining, process mining is described as follows: "The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems."

Process Mining faces event data (i.e., observed behavior) and process models (handmade or discovered automatically) [1]. Mainstream data science approaches like data mining, statistics and machine learning techniques do not consider end-to-end process models [1]. Process mining takes them into account. Compared to process science approaches like mainstream Business Process Management (BPM) ones, that deal with design, execution, control, measurement and optimization of business processes with an emphasise on explicit process models, the focus of process mining is not on process modelling, but on the use of event data [1]. In general, process mining is divided into three different types [1, 13]:

1. Process Discovery:
   As-is process models are generated from event logs.
2. Process Conformance:
   An existing as-planned process model is compared with the as-is model of the same process generated from the event log.
3. Process Enhancement:
   An existing as-planned process model is extended through continuous comparison with actual data from event logs to improve the existing model.

Figure 1 shows how the three process mining types mentioned above and their relations to each other. The process mining types communicate with the real world through process models and interact with the software environment through event log data, which takes on supporting as well as controlling functions towards the real-world processes. Process models and software systems are also connected through further development aspects [1]. Figure 1 also shows the tasks handled in the present case study. The construction company provided data from the supporting software system and knowledge about as-planned processes. The received data was processed and analysed by the authors to gain explicit knowledge about the as-is processes.

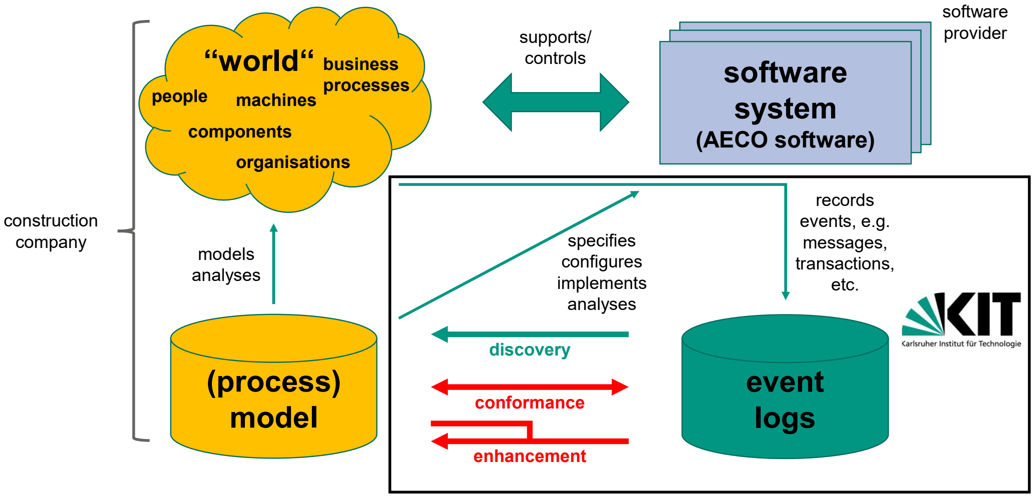

Figure 1. Positioning of the three main types of process mining: discovery, conformance, and enhancement. Allocation of project affiliations. (adapted from [1])

4 Application of Process Mining Methods - Use Case

4.1 Data Collection and Raw Data Export

The German construction company *Wolff & Müller Holding GmbH & Co. KG* uses *RIB iTWO™ 2019 Enterprise* [14] to support processes concerning call for tenders, awarding of contracts as well as accounting and billing. For the process mining analyses in this work, the *RIB iTWO™* data from the bidding phase of a tender of three building construction projects is exported. More precisely, the log data deals with the cost estimation process. Projects 1 and 2 are new construction projects and Project 3 is a renovation project. Because tendering processeses for the three projects were in full progress, no further general information on the projects was processed to ensure data protection. However, for processing and analysing the log data, the elaboration of further general project information was not necessary.

For Project 1 there were three data sets available, each for one individual calculation variant that was conducted by the construction company. Table 1 shows main information regarding the five data sets of the three different projects (Project 1 - Variant 1,2,3; Project 2; Project 3). The exported raw data across projects cover the time span from 14-11-2019 to 04-02-2020. This means that, at the maximum, data is available for a period of slightly more than 11 weeks (Project 1 – Variant 1). In the other data sets, this period is somewhat shorter. This has mainly two reasons. Firstly, a final export of the log data of all projects took place on 04-02-2020. The fact that the date of the last event in four of the five data sets is before 04-02-2020 is because no activities were carried out in the meantime in these projects or project variants. Secondly, for Project 2 and 3 the recording of changes and the log data extraction began one week later than for Project 1 because the realisation of the change logging in the construction company has been slightly delayed for those projects. Since the case study only evaluates a short time frame the exported data of all projects does not cover the entire process for all cases and accordingly all calculated elements. In total, around 100.000 log entries were recorded in all projects. The largest data set for a particular project, namely Project 2, contained about 33.500 entries and thus 33.500 events.

Table 1. Scope of the exported project data sets

Project	First Event	Last Event	Events
1 - Variant 1	14-11-2019 10:48:34	22-01-2020 08:17:54	14.496
1 - Variant 2	14-11-2019 10:48:34	13-01-2020 17:23:01	2.355
1 - Variant 3	14-11-2019 10:48:34	04-02-2020 09:48:26	28.628
2	21-11-2019 09:53:47	27-01-2020 17:45:10	33.495
3	21-11-2019 07:22:40	28-01-2020 17:25:06	20.962

The log entries generally contain information about the executed activity, the user involved, a timestamp and the ordinal number of the processed element, which at the same time represents the case id. An extract of a CSV file exported from RIB iTWO™ is shown in Table 2. One row corresponds to one event. In addition there is a fifth column containing details of the activity carried out. However, since there were already 58 different activities across the projects, this paper only deals with the activities and not with their detailed specifications.

Table 2. Extract from the CSV Log File of one of the Construction Projects (Adopted from Excel)

Timestamp	Ressource (user) - anonymous	Case id	Activity	Activity Detail
03.12.2019 14:35:10	ocsh	1 : 2.20. 1. 1. 90.	OZ: 1 : 2.20. 1. 1. 90. - ME geändert	von "" in "lfm"
03.12.2019 14:35:33	libl	1 : 1.14.12.	OZ: 1 : 1.14.12. - Kurztext geändert	von "Aufzugsmaschinenräume/Unterfahrten Hotel ++ Siehe Bodenbeschichtung" in "Aufzugsmaschinenräume/Unterfahrten Hotel ++ Siehe Bodenanstrich"
03.12.2019 14:35:58	libl	1 : 2.14. 4.	OZ: 1 : 2.14. 4. - OZ geändert	von "" in " 4"
03.12.2019 14:35:58	libl	1 : 2.14.	Gruppenstufe 2.14. 4. in 2.14. eingefügt	
03.12.2019 14:36:47	ocsh	1 : 2.20. 1. 1. 100.	OZ: 1 : 2.20. 1. 1. 100. - Kurztext geändert	von "" in "Reinigungsöffnung"
03.12.2019 14:36:49	ocsh	1 : 2.20. 1. 1. 100.	OZ: 1 : 2.20. 1. 1. 100. - LV-Menge geändert	von "0,000" in "43,000"

Table 3. Data quality issues and R script data pre-processing solution

No.	Issue	Solution
1	Case ids change in event logs due to different types of movement of elements within the bill of quantities	Tracking of changes and assignment of the last ordinal number as unique case id to the related events.
2	Events sometimes do not receive a correct timestamp, as they are partly not automatically logged.	No solution for this quality issue. Further ideas for sorting are to be developed in the future.
3	Some events that belonged to a case were assigned to a parent group of the case within the event log.	The script assigns the entries back to the corresponding case again.
4	A manual activity induces several entries in the change logging, even for other cases.	No solution for this quality issue. The number of events is not to be used as the sole criterion for evaluating the cases.
5	The software logs auxiliary events automatically while executing a virtual intermediate step.	The script eliminates these entries from the event log.
6	After elements are deleted from the bill of quantities, their events remain in the event log.	All events of the associated cases are eliminated by the script.
7	Two different manual activities led to events with the same activity name. It was not possible to determine by the data which activity had taken place.	All events of the associated cases are eliminated by the script.
8	After all the other solution steps there have been still some events in the log data for cases that do not appear in the bill of quantities.	All events of the associated cases are eliminated by the script. However, this observation shows that there are uncertainties in the event log that have not been fully clarified until now.

4.2 Pre-processing the Data

All the five data sets had quality issues which made immediate analysis impossible. The biggest quality issue was the fact that there was no unique case id in the log data. However, the case id and accordingly the ordinal number of an element can change when elements are moved within the bill of quantities in *RIB iTWO^{TM}*. In the data records, like this entries are assigned to different case ids that actually belong together. This issue and others that were detected are listed in Table 3. To address the quality issues the exported raw data is to be processed by a specially created script in the programming language R [15]. It was developed with the help of the open source integrated development environment R Studio [16]. The automated script properly merges all entries that belong to an element by unifying different case ids in related entries. Where applicable, the script also covers the other quality issues. The solutions offered by the script as well as further conclusions from unsolved problems are presented in Table 3. Only through necessary work steps of the script are evaluations with process mining methods possible.

However, it must also be said that the data pre-processing discards over 50.000 events of the original 100.000 events from the raw data. Since process mining analyses depend on the availability of large data sets, discarding such large data parts significantly worsens the general initial situation.

4.3 Discovering the Process Models and further Filtering

The commercial process mining software *Minit^{TM}* [17] was used to execute the process discovery based on the data. However, it is possible to carry out simultaneous analyses with other commercial tools or with open source alternatives such as *ProM* [18]. During this work, different tools were tested equally, but the visual representations of *Minit^{TM}* were subjectively chosen for this paper. *Minit^{TM}* was used with an academic licence.

After pre-processing all five datasets the largest data set is still the one from Project 2. A total of 16.216 events are assigned to 2.918 cases in this data set. Subsequent analyses are presented in this paper representatively based

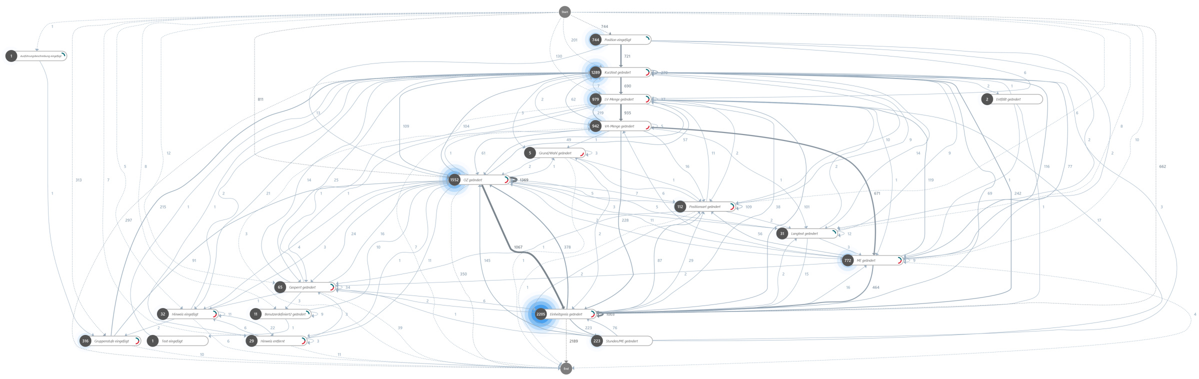

Figure 2. Process model of project no. 2 with 100% of the activities and 100% of the available paths. (adopted from *Minit*^{TM})

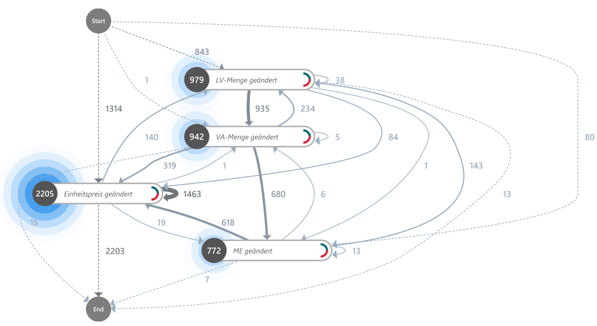

Figure 3. Process model for a data set with four selected activities of project no. 2 with 100% of the activities and 100% of the available paths. (adopted from *Minit*^{TM})

on this data set but could be done analogously to the others as well. Firstly, the control flow perspective is considered. The sequence of the executed activities is shown in Figure 2. It is immediately apparent that the as-is process under study is an unstructured process. One can hardly extract any information from this process model. It is worth noting that Figure 2 accurately reflects the behavior that occurs in the dataset.

In order to obtain a comprehensible control flow model the analysis can be limited to selected activities. In consultation with the lead of process management at the construction company, four activities subjectively considered important were selected for further analysis. Cost estimators were asked for their consent to provide data but they were not further involved in the analysis. For future work it would be beneficial to take the calculators opinion into account too, since they work directly within the software system and have more in-depth operational process knowledge.

Figure 3 shows a process model that only contains the activities "LV-Menge geändert" (engl. "LV quantity changed", as-planned quantity of element), "VA-Menge geändert" (engl. "VA quantity changed", as-is quantity of an element that can change during later stages when project ist carried out), "Einheitspreis geändert" (engl. "Unit price changed", unit price change of an element) and "ME geändert" (engl. "ME changed", change of the unit of measure for an element). After filtering the data set of Project 2 with just the four selected activities the filtered data set consists of 8.855 events and 2.238 cases. The following visualisations and interpretations continue with this filtered data set.

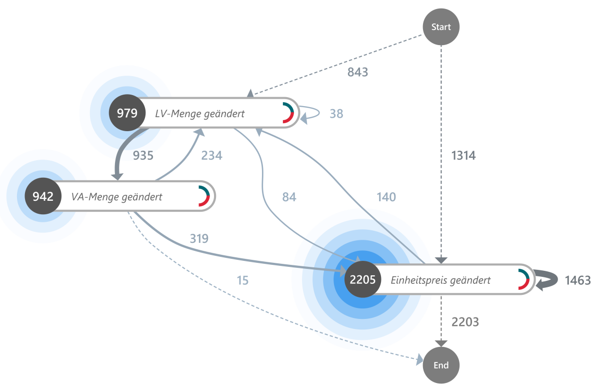

Figure 4. Process model for a data set with four selected activities of project no. 2 with 80% of the activities and 80% of the available paths. (adopted from *Minit*^{TM})

Considering only the most frequent behaviour (e.g. 80% of the behaviour) in the dataset results in the process model in Figure 4. In Figures 2, 3 and 4, the number of cases in which this behaviour was observed is indicated on the activities and the path connection. For example, the activity „LV-Menge geändert“ (change of as-planned quantity of

Table 4. Five cases and accordingly five elements of the bill of quantities where the most activities were carried out from a data set with four selected activities of Project 2 (Adopted from *Minit*^{TM})

CASE	EVENT COUNT	START	END	DURATION
1:1.15.1.1.10. Elektonische Schlosseinheit Clasic offline, Rauch und Brandschutz anforderung VingCard Standard-Zylinder	56	26.11.2019 10:13:06	16.12.2019 16:44:22	20d 6h 31m 16s
1:2.6.1.7.30. Leibungen	47	22.11.2019 08:34:51	16.12.2019 16:44:24	24d 8h 9m 33s
1:1.3.1.15.30. Fllanken Dämmung H= 1m, D= 60mm an der Wand	44	29.11.2019 10:36:26	16.12.2019 16:44:21	17d 6h 7m 55s
1:1.14.3.1.10. Disp. NAK1	39	22.11.2019 11:22:37	16.12.2019 16:44:22	24d 5h 21m 45s
1:2.3.1.4.10. C30/37 Decke bis d = 30 cm	38	29.11.2019 11:30:08	16.12.2019 16:44:23	17d 5h 14m 15s

an element) was carried out for 979 cases and 843 times it was the first activity executed within the process flow.

Even in a simplified process model, as in Figure 4, there are two-way path connections. This means that one cannot derive a directed process flow from this model. In addition, the process model is so simplified that it only reflects the real data set behaviour (see Figure 2) and thus the real process of cost estimation to a limited extent. The timestamps of the entries from the log data are not always in conformity with reality due to the software specific change logging and therefore a consideration of the time perspective is omitted at this point. Hereafter, there will be a view on the data records from the case perspective and the organisational perspective presented.

Table 4 shows an exemplary data set view from the case perspective. It is possible to sort the individual elements according to the number of activities carried out in connection with them or one can read out the respective processing start and end times. Based on this, it is possible to draw conclusions about the effort required to calculate an element. To this aim, however, it is absolutely necessary to look at the individual activities carried out regarding the single elements as well. This is because there are some activities that took place at the same time due to the specific change logging of *RIB iTWO*^{TM}. The absolute number of activities does not correspond to the number of activities carried out in reality. In addition, there are some activities that do not correspond to any real activity. These are activities that reflect a virtual intermediate step that is automatically carried out by *RIB iTWO*^{TM}.

Figure 5 shows an exemplary data set view from the organisational perspective. It is a so-called "social network" [1]. Staff roles can possibly be derived from the connections shown. In Figure 5, "nrti" has the most case involvements. With the exception of "namn", to which there are reciprocal connections, other paths lead unilaterally to "nrti". Presumably, "nrti" takes over controlling tasks in Project 2 and coordinates them with "namn". According to Figure 6, the employee "dabu" carries out activities mostly in a self-loop and there are only outgoing connections to other employees. This can be an indication of the processing of specific crafts or elements, which are then released for further controlling. Such interpretive approaches can be used to highlight valuable information about collaboration within a project. The process manager can see if there are unexpected handovers of tasks, unwanted loops in handovers or missing links between staff. However, it must be checked on the basis of actual role distributions whether such observations correspond to reality. To this aim, data protection-relevant preliminary considerations must be taken into account and the consent of the participating stakeholders for personal data-related investigations must be obtained.

Furthermore, it is possible to combine the case perspective and the organisational perspective in order to gain interesting insights. For this purpose, individual elements can be selected in the case perspective and their respective social networks can be examined. From the combination of case and organisational perspective, correlations can be determined, for example, between cases involving higher efforts and the number or the combination of employees involved in the case.

5 Conclusions and Future Work

Summarising it is necessary to examine the entire cost estimation period in order to be able to map the complete process flow. The recording of the project data must be continuous and linked to the project start and end dates to document the entire project. With such a shortened observation period, like with approximately 11 weeks in the present case study, it is highly probable that a large number of the cases and accordingly the elements from the bill of quantities did not run through the entire cost estimation process to be examined.

Within the scope of the present work, an interface was created between the software *RIB iTWO*^{TM}, that supports the cost estimation processes in a construction company, and common process mining software. A specially developed R script prepared non-continuous case ids which exist in this unsuitable form from the data structure of *RIB*

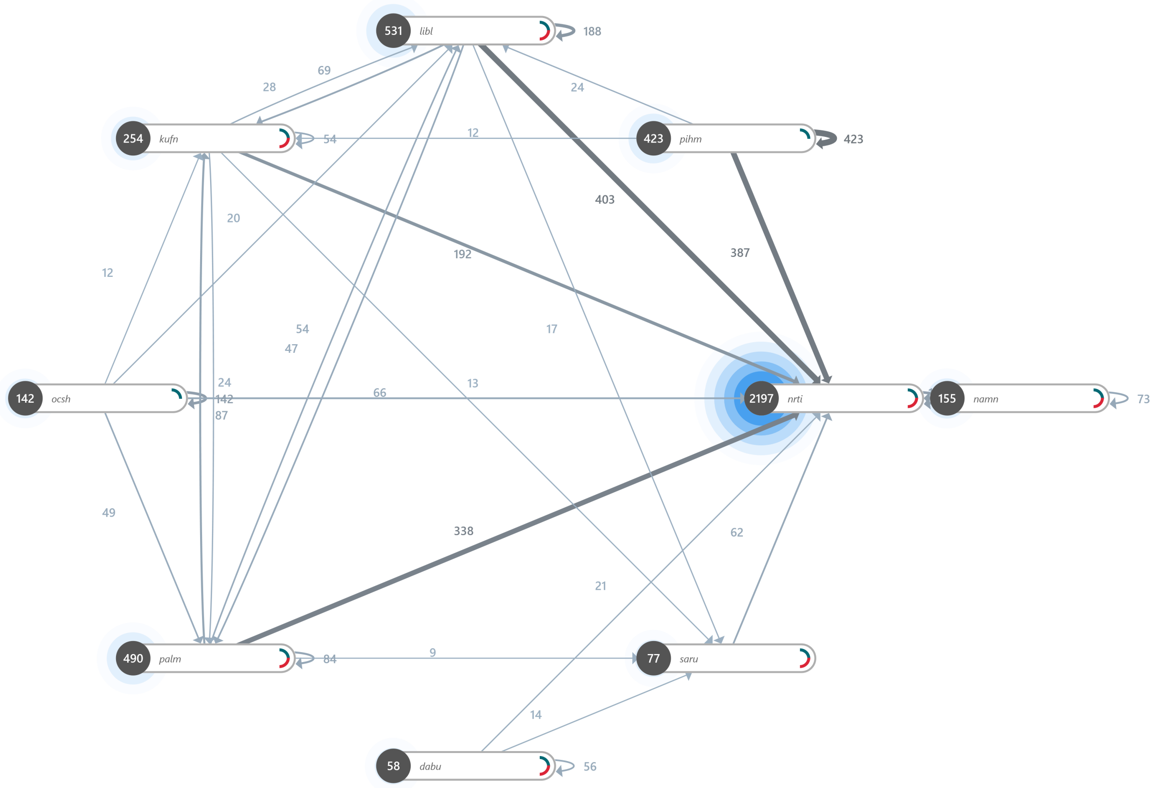

Figure 5. Social network with absolute number of cases and accordingly elements of the bill of quantities worked on. Data set with four selected activities of project no. 2 with 80% of the personnel ressources and 80% of the path connections between the personnel ressources. (adopted from *Minit*^{TM})

iTWO^{TM}. Only by implementing a unique identification number it can be ensured subsequently that elements of the bill of quantities can be traced over the entire cycle of their existence. The case id is the key element in process mining and is absolutely dependent on consistency. By filtering, data with uncertainties can be removed from the event log to enable process mining analysis. However, this eliminates large parts of the event logs which is problematic for well-founded analyses. To prevent the discarding of data in the future, consistent case ids must already be created for elements from the bill of quantities within the utilised software systems.

It has to be mentioned that the generated as-is control flow models from the present case study are on a different and more detailed level of abstraction than the manually created as-planned process models of the construction company which define general process flows in a comprehensible way. This makes process conformance checks or enhancement approaches impossible without further adjustments. To enable further investigations that go beyond process discovery, the level of abstraction from as-is and as-planned models have to be aligned in future work.

At this point, it is advised to look at the event logs from the case perspective, the organisational perspective or a combination of both in order to find starting points for further analysis. If starting points for process optimisations are found by looking at the data from different perspectives, these must be communicated to the employees, that execute the real processes. An examination from the time perspective was not possible with the *RIB iTWO*^{TM} data sets because the timestamps were not always in conformity with reality.

As shown, process mining methods can be used in the context of construction planning projects, when process relevant log data from supporting software systems, like *RIB iTWO*^{TM}, is accessible. The investigations may provide valuable insights into complex AECO disciplines that required particular attention during the process execution. However, the quality of the exported data sets is decisive for the success of the analyses to be carried out. There is great potential for improvement here. The software systems that are used in the AECO industry should be equipped with appropriate documentation mechanisms to routinely meet the process mining requirements. In order to enable reliable analyses, it is sufficient if the documented events of the event logs each contain a case id (for case classification), a timestamp (for time classification) and the executed activity or task. Seamless event logging and taking data protection aspects into account, is a basic requirement for the change logging in software systems and the export and analysis of data sets.

In order to be able to carry out comparative analysis across projects and also across companies, it is urgent to introduce uniform semantics (e.g. for crafts, elements, tasks, processes). For this purpose, generally applicable language rules must be developed for the entire sector, which are then referenced by all utilised software systems. Building on this, it may be possible in future to use process mining routinely and profitably in the AECO industry.

References

[1] W. van der Aalst. *Process Mining: Data Science in Action*, volume 2. Springer Berlin Heidelberg, Heidelberg, Germany, 2016.

[2] HSPI Management Consulting. Process mining: A database of applications (2020 edition). On-line: `https://www.hspi.it/wp-content/uploads/2020/01/HSPI_Process_Mining_Database2020.pdf`, Accessed: 27/04/2021.

[3] G. Girmscheid and C. Motzko. *Kalkulation, Preisbildung und Controlling in der Bauwirtschaft*, volume 2. Springer Vieweg, Heidelberg, Germany, 2013.

[4] K. Rashid and J. Louis. Process discovery and conformance checking in modular construction using rfid and process mining. In *Construction Research Congress 2020: Computer Applications*, pages 640 – 648, Tempe, Arizona, 2020.

[5] S. v. Schaijk and L. v. Berlo. Introducing process mining for aecfm: Three experimental case studies. In *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 11th European Conference on Product and Process Modelling (ECPPM 2016)*, pages 481–488, Limassol, Cyprus, 2016.

[6] L. Zhang, M. Wen, and B. Ashuri. Bim log mining: Measuring design productivity. *Journal of Computing in Civil Engineering*, 32(1):04017071, 2018. doi:10.1061/(ASCE)CP.1943-5487.0000721.

[7] S. Yarmohammadi, R. Pourabolghasem, A. Shirazi, and B. Ashuri. A sequential pattern mining approach to extract information from bim design log files. In *Proceedings of the 33rd International Symposium on Automation and Robotics in Construction (ISARC*, pages 174–181, Auburn, USA, 2016.

[8] L. Zhang and B. Ashuri. Bim log mining: Discovering social networks. *Automation in Construction*, 91:31–43, 2018. doi:10.1016/j.autcon.2018.03.009.

[9] S. Kouhestani and M. Nik-Bakht. Towards level 3 bim process maps with ifc & xes process mining. In *eWork and eBusiness in Architecture, Engineering and Construction. Proceedings of the 12th European Conference on Product and Process Modelling (ECPPM 2018)*, pages 103–111, Copenhagen, Denmark, 2018.

[10] S. Kouhestani and M. Nik-Bakht. Ifc-based process mining for design authoring. *Automation in Construction*, 112:103069, 2020. doi:10.1016/j.autcon.2019.103069.

[11] Autodesk. Revit - multidisciplinary bim software for higher-quality, coordinated designs. On-line: `https://www.autodesk.eu/products/revit/overview`, Accessed: 16/07/2021.

[12] L. Chen, S. Kang, S. Karimidorabati, and C. Haas. Improving the quality of event logs in the construction industry for process mining. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 804–811, Banff, Canada, 2019.

[13] W. van der Aalst et al. Process mining manifesto. In *Business Process Management Workshops. BPM 2011*, pages 169–194, Clermont-Ferrand, France, 2012.

[14] RIB. Ava - itwo. On-line: `https://www.rib-software.com/en/solutions/ava`, Accessed: 18/07/2021.

[15] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

[16] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL `http://www.rstudio.com/`.

[17] Minit j.s.a. Minit. On-line: `https://www.minit.io/`, Accessed: 28/04/2021.

[18] H. Verbeek, J. Buijs, B. van Dongen, and W. van der Aalst. Xes, xesame, and prom 6. In *Information Systems Evolution. CAiSE Forum 2010*, pages 60–75, Hammamet, Tunisia, 2011.