Improving Productive Processes Using a Process Mining Approach

Ricardo Ribeiro, César Analide, and Orlando Belo[(✉)]

School of Engineering, ALGORITMI R&D Center, University of Minho, Campus de Gualtar,
4710-057 Braga, Portugal
obelo@di.uminho.pt

Abstract. Today's companies face great challenges when attempting to quest business markets with their demands on product quality and price. However, when a company maintains high efficiency levels on its productive processes usually it has this challenge quite simplified. The great availability of data we have currently on industry plants provides a very interesting support to face this challenge, when combined with new technologies such as process mining. This paper presents a case study where the very recent process mining techniques were applied to a very particular productive process characterized for its low frequency and heterogeneity. To do this, we made some changes to the "L * life-cycle model" methodology, for applying process mining in the identification of tasks with unsatisfactory performance levels, and analyzing the most relevant and critical aspects that influence it.

Keywords: Business processes management · Process modeling and analysis
BPMN · Business process optimization · Bottleneck points analysis
Process mining · Industry 4.0

1 Introduction

Business markets are quite dynamic. Companies should continuously change their way of being in business, having the need to arrange strong arguments to get a "place at the sun". The pursuit for better efficiency on productive processes raise to companies great challenges, imposing better productive models and new ways of thinking in production activities. However, any company that is able to maintain high levels of efficiency in its productive processes has this challenge largely simplified [1]. In this paper, we will approach a real case study, which provided us a very interesting industrial platform with excellent characteristics for applying process mining technology for studying and optimizing its productive processes. This platform is integrated in a company with worldwide level operations, having several business units in much diversified areas of activity. Spite of this diversity, we will focus our attention on a specific industrial process having very peculiar characteristics that require different methodologies. The process selected has a low frequency. In a typical year of activity, it is triggered by the production of a couple of hundred units. However, all the units produced have frequently a very distinct nature, and a very high production value. The high value of these units is due to three

major factors, namely: (1) the high cost of the raw materials needed to manufacture the unit; (2) the cost of the technology to be applied; and (3) the uniqueness of the characteristics of the unit, which prevents further industrialization of the process. Regardless of the product concerned, it is much more efficient to produce equal units than separate units, and in this particular case the product design process lasts for several months and exceeds the duration of the manufacturing process and that is why we will focus here our optimization efforts. Additionally, the design process is very time-consuming because its development is based on many parameters, which must be carefully verified and validated in a way that minimizes the risk of failure. Failures can lead to the application of guarantees, penalties and lead to a negative impact on the company's reputation in the markets. So, it is very important to prevent any issue that can lead to the occurrence of a failure along the entire production cycle. By nature, the high number of technological parameters involved with product production must be worked by highly skilled labor. The access to specialized labor is very expensive. So, any optimization that occurs on this process will result in great benefits for the company.

In this paper, we will use process mining for identifying tasks with unsatisfactory performance levels approached in the referred productive process, describing and analyzing the most relevant and critical aspects that influence it. In the following sections, we will expose and discuss some related work (Sect. 2), present more detail on the case study used (Sect. 3), and how we design and implemented the process mining (Sect. 4), evaluating and discussing some of the findings (Sect. 5). Finally, in (Sect. 6), we present a brief set of conclusions, pointing out some working lines for future work.

2 Related Work

Companies, regardless of their degree of technological maturity are increasingly having more and more computer systems for supporting their activities, using a diversity of information systems, such as CRM (Customer Relationship Manager) or ERP (Enterprise Resource Planning) systems. They are used directly in the support of productive processes, producing a digital trace that can be used as raw material to be exploited by the very recent process mining (extraction) techniques [2, 3]. Process mining provides us a way for bridging the gap between conventional data mining techniques [4] and process management techniques [5]. Data mining techniques are basically focused on a set of data, usually having little process breadth, while process management techniques are essentially focused on the development of process models that, most of the times, are heavily disassociated with the event records of the process itself, that is, of their instances, which makes any analysis highly unreliable given its strong virtual characteristics. Process mining techniques rely heavily on the information we have about process events and on the quality of the data we have. Both can be used to support a large diversity of tasks, such as discovery [6], compliance verification [7] and enhancement [8] of processes. Discovering knowledge about one or more processes can play an important role in improving or adapting a process model implemented in a specific context. Furthermore, the verification of conformity allows for observing deviations from the imposed models and, consequently, for contributing to the improvement of the

processes. Using knowledge discovery techniques we can also observe many aspects related to the execution of a process, analyzing performance, frequency, resources, occupation and forecasts [3] indicators. All great elements for supporting process management.

Today, we have already access to a set of guidance standards for analyzing performance using process mining [9], which should be followed in order to obtain greater benefits from the application and exploitation of the respective techniques. All the records about the observed events should have the highest possible quality, according to the business requirements and mining parameters defined. Event log data extraction should be question driven, and the designed models should, wherever possible, support competition actions in terms of control of the various flows (data or control) established in the models. These records should relate to the various elements of the process model. The models should be treated as intentional abstractions of reality and the application of any process mining should be carried out continuously. The quality of any model should be assured. We can do this using some well-defined criteria, which considers issues such as robustness, simplicity, precision and generalization. The robustness of the model is evaluated by the ability of the model to map the events in the registers - the more events are mapped, more robust is the model. The simplicity of the model is evaluated by its ability to explain the behavior described in the event logs in the simplest way, with the smallest number of activities. The precision of the model is assessed by the ability of the model to limit its possible variants. Finally, the generalization of the model is evaluated by its capacity to support the various possible behaviors of the process.

Process mining performs analysis over historical data. However, when we have "real-time" event logs available and adequate computing power, it is possible to analyze an instance of a process during its execution time, which allows for us to identify and report conformity in a timely manner. Developing real-time process mining techniques has the potential to increase support for operational decisions that are intrinsically immediate in nature. Even with respect to real time, it makes perfect sense to implement methods and techniques for extracting, transforming and streaming or transmitting ETS (Extract, Transform and Stream) events, or ETT (Extract, Transform and Transmit), which is similar to ETL (Extract, Transform and Load) [10], varying only with respect to the last component, the "Load" that in this particular case must transmit or stream the data properly transformed to standard XES [11] by the appropriate communication channels. These methodologies will certainly be the basis for making possible continuous process mining optimization, which means having process mining always "on".

By establishing a comparison with the so popular OLAP (Online Analytical Processing) [12] tools, it would be possible when implementing ETS or ETT methods and techniques to have OLPMO (Online Process Mining Optimization) tools. In our opinion it can contribute to the materialization of the vision of the 4th industrial revolution [13], particularly at the cognitive level in the proposed 5C architecture for the implementation of physical-cybernetic systems [14]. Finally, we would like to emphasize the potential of combining other types of analysis techniques with process mining, for example data mining associated with artificial intelligence has matured numerous

techniques for analyzing data sets, classification, regression, or grouping, among others, which can be very useful in supporting the predictions [15].

3   Modeling the Process

The productive process targeted represents a significant part of the productive activity of the company. Basically, it has the objective of generating plans for manufacturing. This process has very peculiar characteristics: uniqueness and a low frequency. In order to be possible to use process mining techniques in this case, we made a few changes in the "L * life-cycle model" methodology of process mining [2]. The first change is relative to the uniqueness of the product, focusing in the definition of the grain of the process. Since all the products are potentially unique, an infinite number of distinct processes are required in their production, which made itself the process mining model unfeasible. So, we introduced the concept of "process grain", which can be defined as the lowest level of aggregation of activities that allows for achieving an acceptable variability of processes regardless the product we are dealing with, this can be measured by the process model precision [9]. The second change was related to the low frequency of the process, that is, relative to the small number of instances usually necessary to reach a good coverage of the possible cases. It allows for discovering models with good quality [14]. To help mitigate the effects of this problem, we have chosen to model the process a priori (Fig. 1) to aid the interpretation of the data resulting from the application of process mining techniques in the discovery, verification of conformity and enhancement of the process – note that, for confidential reasons, all the tasks of the process are masked. We know [2] that the existence of manually-modeled processes, even if they have low quality and are disconnected from reality, should be used as much as possible for exploiting existing knowledge, for example existing models can help to define the scope of the process and interpret the data in the event log.
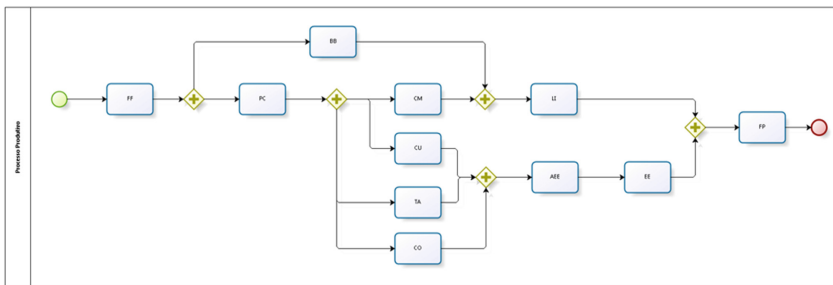

Fig. 1. The BPMN schema of the productive process.

The process we modeled generates the plans for manufacturing the product referred, if we exclude the activity (FF), the receipt of the definition information, and the activity (FP), the conclusion activities, all the rest are related to the different components of the product and all without exception carry out the material reserves in the ERP, issue the plans for manufacturing, and generate the necessary information to produce the next

component. Observing, the process begins with the receipt and blocking of a worksheet (FF) by department 'A'. Whenever necessary, the person in charge of department 'A' meets with the manager of the contract, the issuer (FF), to carry out a more detailed analysis of the requirements of the product. Revisions to the worksheet are carried out in the internal process support system (IPSS), and by the contract manager or by the person responsible for the department issuing the worksheet (FF), which are always accompanied by the person in charge of department 'A'. To ensure that the revisions will be carried out, the worksheet must be unlocked by the department 'A', and after the changes have been made, the worksheet is re-entered and blocked. The issuance of materials requisitions and components to the purchasing department is carried out by the person in charge of department 'A', after checking the consistency of the manufacturing diagram with the deadlines and the workload of the team responsible for the order. The requisitions are carried out at IPSS, which sends the purchase orders to the ERP. In turn, the activity plans (CP) are drawn up after a preliminary study and submitted for approval. In parallel with the activity PC, the activity BB is taking place and once the activity has been completed PC the activities CM, CU, TA and CO start. Once all these activities have been completed, there are two synchronization points that allow the activation of the activities LI and AEE. The activity LI once completed allows for moving directly to the synchronization point that guarantees access to the finishing activity FP, while the activity AEE is still followed by the activity EE before reaching the same point of synchronization, this step allows advancing to the activity for finalizing the process, when it sees all its conditions of performance satisfied.

The activities of the process are registered in different environments. The data came from the IPSS and process management reports. The activity FF is registered in the IPSS, where the beginning of the activity is the record of the availability of the worksheet and the conclusion of the activity is the blocking record of the same worksheet. The activity PC is responsible for issuing plans for client approval. It is registered in the IPSS and the beginning of the activity refers to the date of preparation of the first plan, and the respective conclusion refers to the date of submission for approval of the last plan. The activities BB, CM, CU, TA, CO, LI, AEE and EE are all related to different components of the product, having their sequence due to the dependencies they exert on each other, which are imposed by the current structure of the process. However, these activities, regardless of their distinct nature have a common objective that is the issue of the respective manufacturing plans and, as such, would not expect another criterion other than this to mark their actions in the process. This means that the beginning of the activity is associated to the beginning of the calculations necessary for the component dimensioning, and the conclusion of the activity is associated to the availability of the respective plans for manufacturing. The records of these activities exist in the IPSS associated to the plans issued, and there are also in the ERP records associated with the manufacturing orders issued. Lastly, the activity FP, an activity that can be classified as control and validation, reveals its existence only in the process management report, where its working times have been collected. Given the nature of the systems that support this process, and since there are several conditioning factors for collecting the times, they were all corroborated by the process management reports for minimize errors in data extraction. Regardless of the diverse nature of the information, it was possible to extract

and normalize the data to obtain the set of process events records that allows the use of process mining techniques.

4   The Process Mining

To be considered useful, the registration of an event must have information related to the project to which the activity concerns. In this case the project is the aggregator of activities, the unique number of the activity identifier, a short name of the activity, the date and time of activity start, date and time of activity completion, abbreviated name of the resource that performed the activity, and finally the role that the resource represents in the organization. On the left side of Fig. 2, it is possible to see a subset of the data that was collected, and on the right side, its conversion to the XES (Extensible Event Stream) standard [11]. This standard published by the IEEE (Institute of Electrical and Electronics Engineers) aims at interoperability in registers and event flows. This conversion is useful for normalizing the data [16]. However, it is also necessary because it allows for feeding several process mining algorithms residing in the tool that we will use in the analysis: ProM [17]. This tool is a project of the process mining group of the University of Eindhoven [18], aiming to provide a framework for supporting process mining algorithms. The tool is free and open source. ProM provides an extension built specifically to aid in converting data to the XES standard. Automatically, as we provide the tool with the location of our data file, and in case it is not already in accordance with the standard, this automatically suggests the use of the import extension. At this stage, it only remains validating some parameters of the converter and performing the necessary mapping so that the extension can correctly perform the conversion of the data to the XES standard. This conversion allows the use of other extensions specifically built for performing the extraction and analysis of information about the process.

Fig. 2. Event records of the process activities converted to the XES 1.0 format.

Process information is presented after a successful conversion and is divided in three areas, namely control, inspection and summary. The control information details the number of processes that were imported (1 process), the number of instances of the processes (45 instances), the number of event classes of the process (22 classes of events), the number of events per instance (22 events), the number of event classes per instance (22 classes of events), and the time interval contemplated by the event register (start - October 9, 2012, end - September 13, 2013). The inspection information details instances, events of the instances and their respective attributes, and the frequency of

events. The summary information details the number of process instances, the number of events recorded, the classes of events with their absolute and relative occurrence, and finally the instance and start event classes. With the data presented, it is possible to obtain an indicator that characterizes the quality of the changes that we promoted on the "L-life-cycle model" methodology. Analyzing the number of event classes of the process and the number of event classes per instance, we easily notice that the frequency of event classes is high throughout all instances of the event log, which means that the mode precision is also high. At this stage, it is possible to state that the grain of the process was correctly defined and applied, which allows for obtaining a homogeneous process with little variability and enabling the following analysis. After importing data, a check on the distribution of the process instances in time, and of their durations is a must for detecting gross errors that may have occurred in data extraction, transformation or loading. This checks are intended to gauge the integrity of the data, which is one of the most important factors for the success of a process mining [16]. An instance of the process is only consistent if and only if there is a one-to-one correspondence between its start and end events. After the data integrity and validation phases, we applied the process discovery algorithms on our event log. The discovered process model can be seen in Fig. 3. Several algorithms can be applied for discovering process models [19–26]. However, all these alternatives present different behaviors, limitations and results so the criterion in their selection is decisive for the success in the discovery and the subsequent analysis. The algorithm that we selected [27] takes into account some properties of our data, particularly the ones with respect to the information related to the activities life cycle transitions - "start" and "complete" (Fig. 2). The algorithm we used, in addition to guaranteeing the return of a robust model [28] has the ability for distinguishing concurrency and interleaving. Although the designed model did not identify the interleaving of activities, the reality of the data could reveal a different scenario.
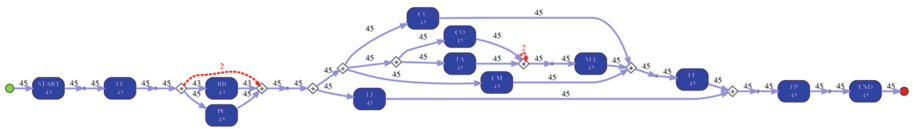

Fig. 3. The discovered process model.

The uncovered model (Fig. 3) has 13 activities, two activities increase on the 11 activities observable in the model we designed manually (Fig. 1). This is so because we used an extension of the tool specifically created for the effect to artificially add these two activities - "START" and "END"- to the model, which were added only to guarantee the correct operation of the algorithms of performance analysis. The activity performance analysis significantly supports the evaluation of the model and assists the validation of the optimization actions. In Table 1 we can see a compilation of the model performance, where we can find the number of occurrences, the number of deviations, the waiting time (the time an activity became available and (the time a resource is busy with an activity), and the total time (the sum of the waiting and execution times plus the

synchronization time, which is for concurrent activities the time between the completion of the first and the conclusion of the last).

Table 1. Discovered model performance results.

Activity	Instances	Deviations	Waiting times	Execution times	Total times
START	45	0	0d 00:00:00:000	0d 00:00:00:000	0d 00:00:00:000
FF	45	0	0d 00:00:00:001	1d 12:49:08:000	1d 12:49:08:001
BB	43	2	3d 12:52:44:000	10d 08:59:06:976	13d 21:02:13:953
PC	45	0	3d 05:18:34:666	21d 15:49:21:333	24d 21:07:56:000
CU	45	0	3d 20:58:06:976	10d 08:12:43:255	14d 05:10:50:232
CO	45	0	4d 01:54:01:428	2d 22:02:57:142	6d 21:43:30:697
TA	45	0	3d 21:06:48:571	7d 09:36:15:714	11d 05:23:23:720
AEE	45	0	6d 05:09:32:000	5d 20:40:41:333	12d 01:50:13:333
CM	45	0	4d 03:36:20:000	6d 03:38:31:428	10d 05:57:53:023
EE	45	0	3d 01:05:38:666	10d 12:44:06:666	13d 13:49:45:333
LI	45	0	13d 18:19:47:441	18d 05:36:41:860	31d 23:56:29:302
FP	45	0	3d 01:09:32:000	2d 14:17:46:666	5d 15:27:18:666
END	45	0	0d 00:00:00:000	0d 00:00:00:000	0d 00:00:00:000

5   Analysis of Results

Excluding the artificial activities, the process discovered differs a little bit from the one we draw previously. Never the less, we analyzed these differences, as they may be key elements of a potential optimization. The execution of the activity BB differs in comparison to that of the model drawn in the synchronization point. In the discovered model, it is synchronized with the activity PC, and in the model drawn with the activity CM. This difference indicates that in the universe of analyzed data the activity BB has no dependencies from the activity CM. The activities CU and CM also differ from the model drawn at the point of synchronization. In this case, the synchronization point is performed after the activity execution (AEE), which can be an indicator that we are in the presence of a dependency or that this is a situation caused by resource management. These are the differences that we register between the two models. However, the discovered model provides some more information regarding deviations (Fig. 3) that occur to it. The deviations are indicating that there are two instances of the process that do not conform to the model. The activity BB for these two instances starts as modeled, but does not respect the point of synchronization with PC, synchronizing only at the point of synchronization of CO and TA. These deviations can reveal different situations, which may be related to the heterogeneity of the product, with particular logistics situations, with rework, or, even with resource management. Anyway, and since its frequency is seen in less than 5% of the instances we are going to treat these cases as noise. Analyzing the results presented in (Table 1), and even for those without a broad knowledge of the business, we easily perceive that there are two activities (PC and LI) that strangulate the flow of this process this can be understood only by observing the high total times. There

is also an indicator that points to the need for optimization - this indicator is the high waiting times associated with practically all activities, but with special emphasis on the activity LI, and that we have previously seen to be a bottleneck of the process, here we quickly find that the resources of the process are over-allocated, which may be due to the volume of work or planning efficiency. These identifications are critical in an optimization process because they allow for focusing the actions in the most problematic areas.

6   Conclusions and Future Work

We showed that process mining techniques can answer multiple questions and produce suggestions for process improvement, even for cases where processes have low frequencies and heterogeneous natures. The improvements can be greater as more data sources become available for improving the dataset. Some of the data sources available were not used mainly due to the delay revealed in obtaining the access to the data, i.e. we had the confirmation of existing data on the ERP concerning the creation of the elements for the product bill of materials, but we could not obtain it in time to include it in this work. However, it is important to investigate and develop some techniques or methods of process mining that are suitable for the analysis of heterogeneous processes. Otherwise, they will always be difficult to analyze. We started this work with suggesting some changes to the "L-life-cycle model". Without them, we believe it would be impracticable to apply the techniques we used. At the short term, we intend to start the construction of a production system and its respective data warehouse, as well as the use of artificial intelligence techniques in the construction of a model for predicting the duration of activities based on product and process attributes. This will aid in the predictions of the behavior of future instances in the model. These actions should create the conditions to allow for having online process mining with optimization OLPMO and thus contributing also to a small part of the vision on the 4th industrial revolution.

References

1. Utterback, J.: Mastering the dynamics of innovation: how companies can seize opportunities in the face of technological change. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship (1994)
2. van der Aalst, W.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer Science & Business Media, Heidelberg (2011)
3. van der Aalst, W.: Process mining in the large: a tutorial. In: LNBIP, vol. 172, pp. 33–76 (2014)
4. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Elsevier/Morgan Kaufmann, Waltham (2012)

5. Van Der Aalst, W.M.P., Ter Hofstede, A.H.M., Weske, M.: Business process management: a survey. In: International Conference on Business Process Management, pp. 1–12 (2003)

6. Van Dongen, B.F., Alves De Medeiros, A.K., Wen, L.: Process mining: overview and outlook of Petri net discovery algorithms. In: Jensen, K., van der Aalst, W.M.P. (eds.) LNCS, pp. 225–242. Springer, Heidelberg (2009)

7. Van Der Aalst, W., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. **2**, 182–192 (2012)

8. Van der Aalst, W.M.P., Schonenberg, M.H., Song, M.: Time prediction based on process mining. Inf. Syst. **36**, 450–475 (2011)

9. Van der Aalst, W.: Process mining: overview and opportunities. ACM Trans. Manag. Inf. Syst. **3**, 7 (2012)

10. Kimball, R., Caserta, J.: The Data Warehouse ETL Toolkit (2015)

11. IEEE Std 1849-2016: IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. IEEE (2016)

12. Chaudhuri, S., Dayal, U.: An overview of data warehousing and OLAP technology. ACM SIGMOD Rec. **26**, 65–74 (1997)

13. Lasi, H., Fettke, P., Kemper, H.G., et al.: Industry 4.0. Bus. Inf. Syst. Eng. **6**, 239–242 (2014)

14. Lee, J., Bagheri, B., Kao, H.A.: A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manuf. Lett. **3**, 18–23 (2015)

15. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)

16. van der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., et al.: Process mining manifesto. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) pp. 169–194. Springer, Heidelberg (2011)

17. van Dongen, B., de Medeiros A.K.A., Verbeek, H.M.W. et al.: The ProM framework: a new era in process mining tool support (2005)

18. Process Mining http://www.processmining.org/start, http://www.processmining.org/. Accessed 9 Sep 2017

19. Van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Trans. Knowl. Data Eng. **16**, 1128–1142 (2004)

20. Redlich, D., Molka, T., Gilani, W., et al.: Constructs competition miner: process control-flow discovery of BP-domain constructs. In: LNCS, pp. 134–150 (2014)

21. Redlich, D., Molka, T., Gilani, W., et al.: Scalable dynamic business process discovery with the constructs competition miner. In: CEUR Workshop Proceedings, pp. 91–107 (2014)

22. Schimm, G.: Process miner - a tool for mining process schemes from event-based data. In: LNCS, pp. 525–528 (2002)

23. Burattin, A., Sperduti, A.: Heuristics miner for time intervals. In: Proceedings of the 18th European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning, ESANN 2010, pp. 41–46 (2010)

24. Günther, C.W., Rozinat, A.: Disco: discover your processes. In: CEUR Workshop Proceedings, pp. 40–44 (2012)

25. Wen, L., Wang, J., Van Der Aalst, W.M.P., et al.: A novel approach for process mining based on event types. J. Intell. Inf. Syst. **32**, 163–190 (2009)

26. Schimm, G.: Mining exact models of concurrent workflows. Comput. Ind. **53**, 265–281 (2004)

27. Leemans, S.J.J., Fahland, D., Van der Aalst, W.M.P.: Using life cycle information in process discovery. pp. 1–12 (2016)

28. Van Der Aalst, W.M.P., Van Hee, K.M., Ter Hofstede, A.H.M., et al.: Soundness of workflow nets: classification, decidability, and analysis. Form. Asp. Comput. **23**, 333–363 (2011)