Analyzing interconnected processes: using object-centric process mining to analyze procurement processes

Alessandro Berti[1] · Urszula Jessen[2] · Gyunam Park[1] · Majid Rafiei[1] · Wil M. P. van der Aalst[1]

Abstract

The purchase-to-pay (P2P) process is one of the core business processes in any organization. It ensures the correct and efficient provisioning of materials and services. An efficient P2P process reduces operational costs by ensuring discounts, avoiding late payments, and choosing the optimal supplier for the goods. Process mining techniques help practitioners optimize the execution of P2P processes by analyzing the execution data and providing useful insights. However, existing techniques may result in misleading insights due to many-to-many relationships between business objects, e.g., between orders and invoices in the P2P process. Recently, object-centric process mining techniques have been proposed to avoid the limitations of traditional process mining techniques. In this paper, we present a case study on a real-life P2P process using object-centric process mining techniques. To that end, we adopt the well-known PM^2 process mining project methodology to the object-centric setting and analyze the performance and compliance.

Keywords Object-centric process mining · SAP ERP · Purchase-to-pay

1 Introduction

Process mining is a branch of data science aiming to exploit the event data recorded during the execution of an organizational process to get insights into the process. In particular, process discovery (the automatic discovery of a process model from the event data), conformance checking (the comparison between the behavior recorded in the event log and the process model), model enhancement (enriching the process model with frequency/performance information) and

predictive analytics (predicting the remaining time or the next activity of an incomplete case) techniques have been proposed. Organizations successfully applying process mining (e.g., Siemens, BMW, Lufthansa, Uber, and Zalando) have reached a scale allowing them to save millions of Euros Reinkemeyer [15].

To apply process mining successfully, the process mining discipline provides several project methodologies aiming at supporting the application of process mining in organizational contexts. For instance, the L* life-cycle model [18] and the Process Mining Project Methodology (PM^2). van Eck et al. [20] provide clear guidance to practitioners on how they implement process mining projects, which aim to improve process performance and compliance to rules and regulations.

A potentially deleterious assumption of process mining is the association between events and cases/process instances. In particular, it is assumed that an event is associated with a single case. For example, the event of resolution of a ticket is associated with the ticket case in a ticketing management system. However, this is unrealistic in many real-life scenarios. Considering a purchase-to-pay process (P2P), a purchase order can be associated with several invoices and each one of which can be associated with different payments. Vice versa, an invoice can be associated with different purchase orders.

✉ Alessandro Berti
  a.berti@pads.rwth-aachen.de

  Urszula Jessen
  urszula.jessen@ece.com

  Gyunam Park
  gnpark@pads.rwth-aachen.de

  Majid Rafiei
  majid.rafiei@pads.rwth-aachen.de

  Wil M. P. van der Aalst
  wvdaalst@pads.rwth-aachen.de

1 Process and Data Science Group, RWTH Aachen University, Ahornstrasse 55, 52074 Aachen, NRW, Germany

2 Process Insights, ECE Group Services, Heegbarg 30, Hamburg, Hamburg 22391, Germany

Requiring events to be associated with a single case may lead to *deficiency* (events not associated with any case), *convergence* (an event needs to be replicated for different cases), and *divergence* (several instances of the same activity are contained in the same case) issues explained in van der Aalst [22].

Object-centric event logs relax the assumption that an event is related to a single case. Instead, an event can be associated with different objects of different object types (e.g., an order and two invoices). This helps to resolve the deficiency/convergence/divergence issues since different objects can be related to an event, and we do not need to "coerce" events inside a case notion (for example, an event of invoice creation does not need to be repeated for all the purchase orders). The discipline of object-centric process mining, i.e., exploiting the information contained in object-centric event logs to obtain useful insights, is in active development.

In this paper, we want to propose a case study of object-centric process mining on top of a purchase-to-pay (P2P) process (Fig. 2 contains a summary of the main stages of the P2P process), along with the description of novel ad hoc techniques that are needed for the analysis. This is motivated by the difficulties in getting reliable insights out of traditional P2P event logs due to convergence/divergence issues. To apply object-centric process mining in an organizational setting, we extend PM^2 to guide the organization that seeks to apply object-centric process mining based on ERP systems data. The renewed methodology consists of six stages and is summarized in Fig. 1.

1. *Planning*: this stage is to set up the project and determine the analysis' goals that need to be answered at the end of the project in a way that improves the process performance.
2. *Extraction*: this stage is to extract the event data from the information system and obtain the object-centric event log.
3. *Data preprocessing*: this stage is to prepare the event data so that the following mining and analysis techniques can produce optimal results.
4. *Mining and analysis*: this stage is to apply object-centric process mining techniques to the preprocessed event data and get insights into the processes which answer the analysis' goals.
5. *Evaluation*: in this stage, the previous stage's findings are validated by the domain experts and interpreted to identify possible action points.
6. *Improvement*: this stage transforms actionable insights into actual management actions that support the process to improve performance and compliance.

In particular, the extraction, preprocessing, and analysis steps required adaptation, since an object-centric event log is extracted (different design choices are possible on the correlation between events and objects of the log) and traditional preprocessing/analysis techniques cannot be applied to an object-centric event log.

The rest of the paper is organized as follows: In Sect. 2, we present the context in which the analysis was performed. In Sect. 3, we plan the case study by describing the process and providing the analysis' goals. In Sect. 4, the extraction of the object-centric event log for the given process is described. In Sect. 5, the preprocessing strategies are presented. In Sect. 6, graph-based and statistical techniques are presented to respond to the initial goals. Section 7 presents the OCPM tool used throughout the case study. In Sect. 8, the results of the analysis are validated and evaluated. In Sect. 9, some adopted and planned improvement strategies for the organizational P2P process are discussed. Section 10 presents the related scientific work. Finally, Sect. 11 concludes the paper.

2 Context

In this section, the context of the company behind the proposed case study is described, along with the initial results obtained from the application of process mining techniques/tools. Moreover, the description of the process is included.

2.1 Company

The ECE Group was founded in 1965 and has since grown to become a leading player in the European shopping center industry. ECE's current construction and planning activities amount to €3.2 billion, and it manages a portfolio of 200 shopping centers. With assets under management of €31 billion and a workforce of 3,300 employees, the company has a strong presence in 13 countries.

In early 2020, ECE Group Services started a process mining initiative aimed at improving its complex and data-intensive processes. The pilot project showed the viability of using process mining tools and methods for process enhancement but also uncovered the need for a structured governance framework and standard operating procedures to conduct effective process analysis. Based on the preliminary results, a dedicated team—*process insights*—within ECE began developing new projects to support various business users in optimizing their processes.

In the early stages of the process insights project, the team utilized a method based on the standard BPM framework, which included process mining tools and techniques, with the aim to enhance business processes. The BPM lifecycle model, widely recognized as the standard, encompasses six essential steps: process identification, discovery, analy-
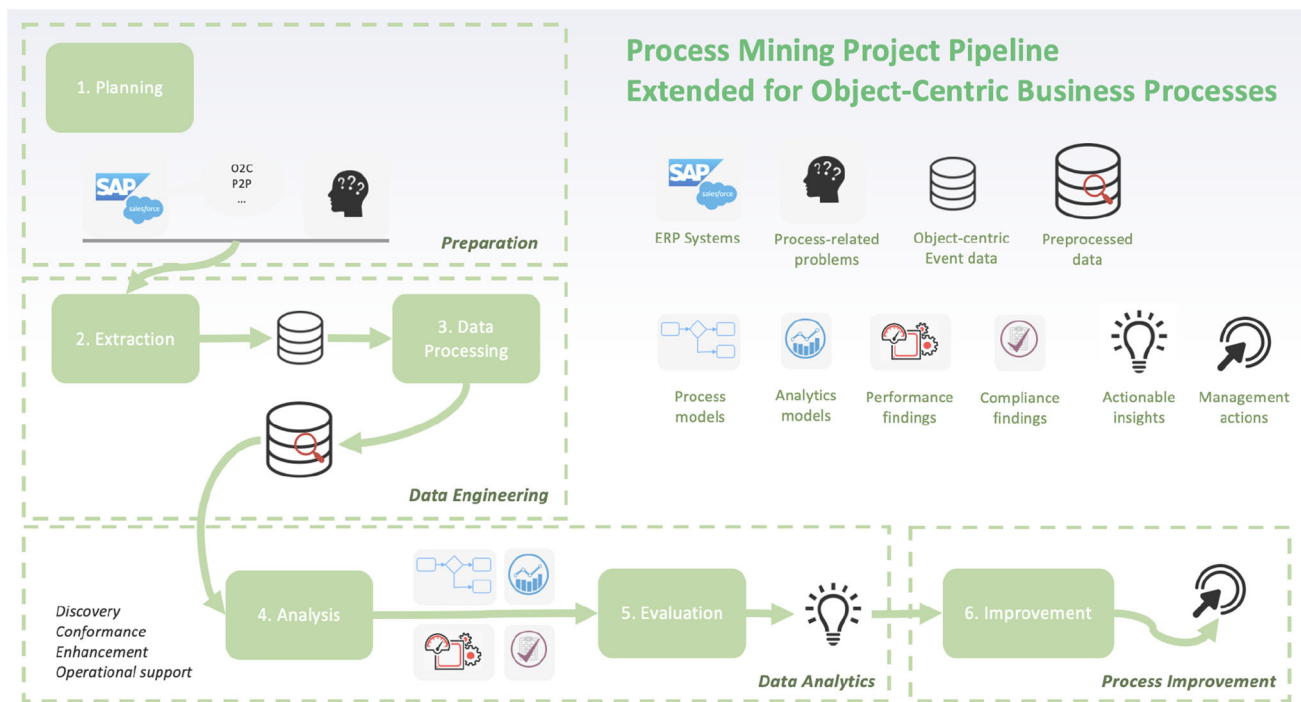
Fig. 1 Adaptation to object-centric process mining of the PM^2 methodology

Fig. 2 Main stages of the purchase-to-pay (P2P) process. The purchasing part includes the management of the purchase requisition, the purchase order, and the receipt of the goods. The payment part includes the verification of the invoice and the subsequent payment

sis, redesign, implementation, and monitoring and control, all aimed at optimizing and streamlining processes [23]. The team's approach in the ECE context involved defining clear goals and objectives, gathering and preparing data, analyzing processes, identifying improvement opportunities, implementing changes, monitoring progress, and continuously re-evaluating processes to ensure desired results are achieved and new optimization prospects are explored.

However, process analysis projects faced several challenges, including a large number of process variants and standards for different countries and stakeholders. With different requirements, the number of KPIs and analyses has been increasing, which made the close monitoring of processes more difficult and time-consuming. Given these difficulties, it is important to obtain reliable and explainable results from the process mining analyses and to identify areas with significant improvement potential, in order to justify the improvement initiatives.

2.2 Initial results and tools

Given the complexity of the aforementioned issues, we initially chose to concentrate on a single process and its interactions with other processes. We began with the *Accounts Payable* process, and, after consultations with top management, the research objective was established as reducing the number of overdue payments. This focus allowed us to effectively identify the root causes of the problem and implement targeted solutions while also laying the foundation for future process improvement initiatives. By focusing on a specific process, we aimed to demonstrate the practical application of process mining techniques and the benefits they can bring to the organization. The objective of reducing late payments was selected as it had a direct impact on the financial performance of the organization and could be easily measured. By successfully addressing this issue, we hoped to demonstrate the value of process mining and encourage wider adoption within the organization.

In the process improvement initiative at ECE, Celonis has been used as the primary tool for analyzing and visualiz-
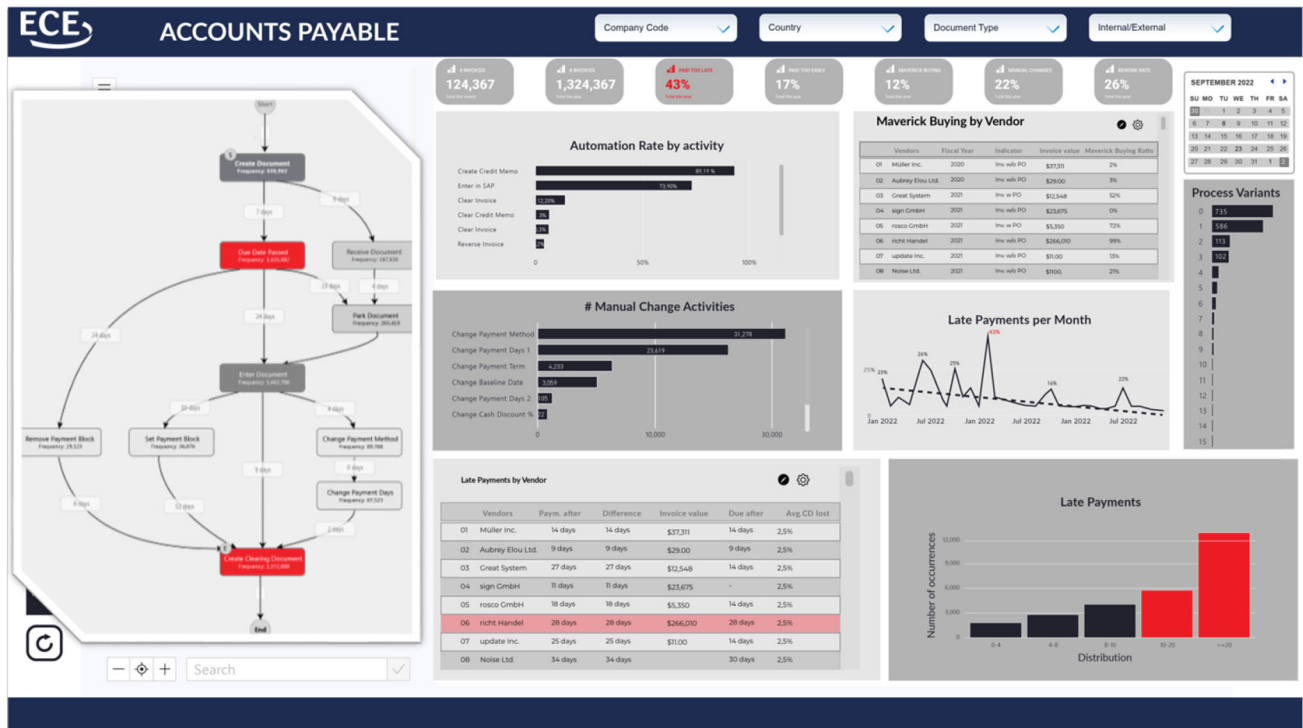
Fig. 3 Simplified representation of a typical process mining dashboard for the Accounts Payable process (which is interconnected with the P2P process)

ing the behavior of various processes. The Celonis Action Engine, a web-based application, was employed to turn process analysis insights into actionable recommendations and to provide operational support during the process execution [2]. Also, process mining dashboards were adopted. Figure 3 shows a simplified and anonymized representation of a typical process mining dashboard for the *Accounts Payable* process. Similar dashboards were used by both process participants and management to optimize the process. It highlights how addressing the different needs and requirements of different stakeholders can result in insights that may be difficult to uncover or put into action.

However, the effectiveness of the tool was highly dependent on the process analyst and how they designed the signals extracted from the data model. It failed to address several challenges faced by the ECE team, such as identifying complex patterns in interconnected processes and providing root cause insights based on feature importance calculations. To overcome these limitations, we decided to form an informal network of experts from various fields and work together to find the best solutions for specific challenges.

Currently, the focus is extended to the purchase-to-pay (P2P) process, which involves both the procurement/purchasing and a payment part (see Fig. 2), and is therefore interconnected with the Accounts Payable (AP) process.

2.3 Process description

The considered P2P process involves the following steps:

1. *Purchase Requisition*: A purchase requisition is created by a user in the SAP system (with the execution of the transaction *ME51N* in SAP) to request the purchase of goods or services. This requisition is then reviewed and approved by a designated approver.
2. *Purchase Order*: Once a vendor has been selected, a purchase order (PO) is created in the SAP system to place the order with the vendor (with the execution of the transaction *ME51N* in SAP). The PO includes the details of the goods or services being purchased, the price, and the delivery terms.
3. *Goods Receipt*: Once the goods or services have been delivered, a goods receipt is recorded in the SAP system to acknowledge receipt of the items (with the execution of the transaction *ME21N* in SAP).
4. *Invoice Verification*: The vendor sends an invoice to the purchaser for the goods or services provided. The invoice is then reviewed and verified in the SAP system to ensure that the goods or services have been received and that the charges are correct (with the execution of the transaction *MIRO* in SAP).

Fig. 4 Execution of the P2P process in ECE. The view includes the activities of the process, the tasks, and the flow of documents and materials
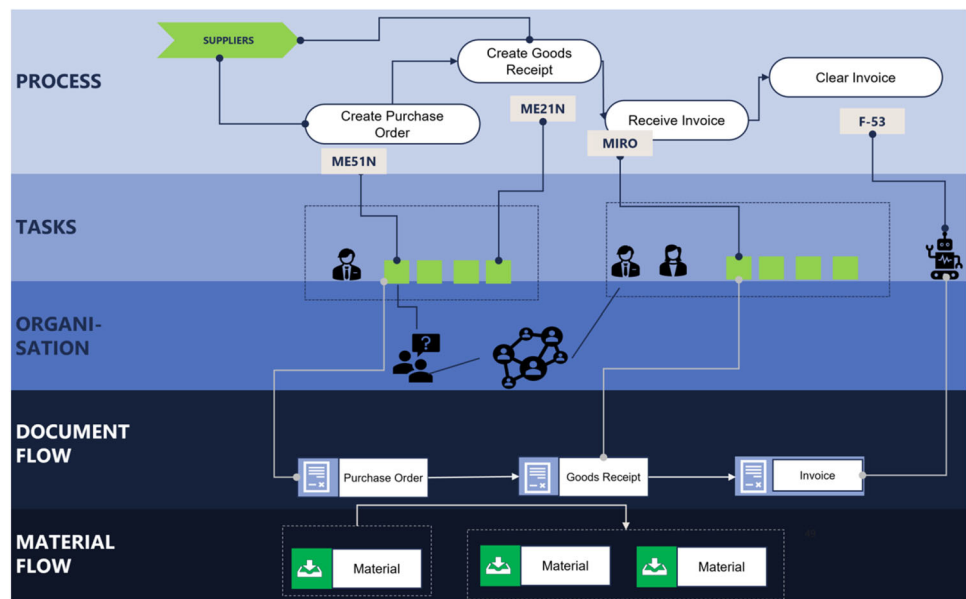
5. *Payment*: Once the invoice has been verified, the payment is processed in the SAP system (with the execution of the transaction *F-53* in SAP). This may involve creating a payment request, issuing a check, or making an electronic payment to the vendor.

The execution of the process is summarized in Fig. 4. The process has been partly automated using the *Invoice xSuite* solution (https://www.xsuite.com/software/invoice/) for the digital acquisition of the documents (purchase requisitions, purchase orders, goods receipts, invoices). The meta-data that is acquired from the software needs to be checked by an operator.

In our research, we analyzed issues pertaining not only to the purchase-to-pay cycle but also to interconnected processes, most notably, the Accounts Payable (AP) process. The AP process is responsible for managing the payments to vendors and suppliers for goods and services that were procured through the P2P process. The steps in the P2P process that are connected to the AP process include:

- *Purchase Order*: This step in the P2P process is connected to the AP process as it involves creating the purchase order and including the details of the goods or services being purchased, the price, and the delivery terms. This information will be used in the invoice verification process and also in the payment process, to ensure that the right vendor is paid the right amount for the right goods and services.
- *Invoice Verification*: This step in the P2P process is closely connected to the AP process as it involves reviewing and verifying the vendor invoices to ensure that the

charges are correct and that the goods or services have been received.
- *Payment*: This step in the P2P process is directly connected to the AP process as it involves processing the payments to vendors and suppliers through the SAP system. The AP process would be responsible for creating the payment request, issuing the check, or making an electronic payment.

Overall, the P2P process in SAP helps to ensure that the right goods and services are procured from the right vendors at the right price, while the AP process in SAP helps to ensure that the vendors are paid correctly and on time for the goods and services that were procured through the P2P process. In light of the aforementioned points, an optimal implementation of the P2P process requires not only a good execution of the procurement and accounts payable steps but also an efficient collaboration between the two stages.

3 Planning

In this section, we plan the case study. In Sect. 3.1, we propose object-centric event logs as the target data format. In Sect. 3.2, the analysis' goals are proposed.

3.1 Target data format

Object-centric process mining techniques require object-centric event logs (OCELs, Ghahfarokhi et al. [12]).[1] An example of OCEL in tabular form is contained in Table 1.

---

[1] http://www.ocel-standard.org/.

Table 1 Example object-centric event log of a P2P process represented as a table

Ev.ID	Activity	Timestamp	Purch.Req.	Purch.Ord.	Goods Issues	Invoices	Payments
e1	Create Purchase Requisition	2021-03-20 10:30	['PR1']				
e2	Close Purchase Requisition	2021-03-20 14:00	['PR1']				
e3	Create Purchase Requisition	2021-03-21 09:30	['PR2']				
e4	Create Purchase Order	2021-03-22 14:59	['PR2']	['PO1']			
e5	Invoice Receipt	2021-03-25 11:00		['PO1']		['R1']	
e6	Perform Payment	2021-03-30 11:58				['R1']	['P1']
e7	Create Purchase Requisition	2021-04-01 09:15	['PR3']				
e8	PR Formal Approval	2021-04-01 10:15	['PR3']				
e9	Create Purchase Order	2021-04-02 17:00	['PR3']	['PO2']			
e10	Change Purchase Requisition	2021-04-03 10:00	['PR3']				
e11	Invoice Receipt	2021-04-05 15:00		['PO2']		['R2']	
e12	Perform Payment	2021-04-15 09:27				['R2']	['P2']
e13	Create Purchase Order	2021-04-17 14:29		['PO3']			
e14	Invoice Receipt	2021-04-28 10:00		['PO3']		['R3']	
e15	Perform Payment	2021-04-30 15:00				['R3']	['P3']
e16	Invoice Receipt	2021-05-28 10:01		['PO3']		['R4']	
e17	Perform Payment	2021-05-30 15:17				['R4']	['P4']
e18	Invoice Receipt	2021-06-28 10:01		['PO3']		['R5']	
e19	Perform Payment	2021-06-30 15:29				['R5']	['P5']
e20	Create Purchase Requisition	2021-07-01 11:15	['PR4']				
e21	Create Purchase Order	2021-07-02 09:38	['PR4']	['PO4']			
e22	Invoice Receipt	2021-07-09 16:00		['PO4']		['R6']	
e23	Goods Issue	2021-07-11 10:30		['PO4']	['GI1']		
e24	Perform Payment	2022-05-15 09:00				['R6']	['P6']
e25	Invoice Receipt	2022-05-20 12:00				['R7']	
e26	Create Purchase Order	2022-05-20 15:00		['PO5']		['R7']	
e27	Create Purchase Order	2022-06-01 09:17		['PO6']			
e28	Create Purchase Order	2022-06-02 11:48		['PO7']			
e29	Create Invoice	2022-06-05 09:00		['PO6', 'PO7']		['R8']	

For example, the first row contains the event with identifier *e1*, activity *Create Purchase Requisition*, and timestamp *2021-03-20 10:30*. This event is related to a single object *PR1* of type *Purch.Req.*. We adopted the JSON-OCEL implementation, which is based on the JSON format.

Generating an object-centric event log from a database requires:

- The extraction of the set business objects, along with their attributes.
- The extraction of the set of events.
- The correlation between events and objects.

In particular, the last step requires a significant amount of design choices. For example, in a P2P process, an order can be related to different invoices. The correlation between orders and invoices can be described in the event with the activity "Create Purchase Order" (which will report the order and all the related invoices) or the event with the activity "Create Invoice" (reporting the order and the invoice related to the event).

3.2 Analysis' goals

We categorize different goals to identify the problems that can be addressed by object-centric process mining. Note that fulfilling these goals requires dealing with interdependencies between different processes, where conventional process mining approaches do not provide convincing answers. We categorize the identified goals into three main categories: Process Performance (PP), Process Compliance (PC), and Process Quality (PQ). The goals related to *Process Performance* (PP) are as follows:

PP1 What is the current processing capacity of the accounts payable department?

PP2 How much time is needed from the accounts payable department to process/verify a single invoice document?

PP3 How many procurement orders were inserted correctly during the creation so that no change is needed? (no-touch orders)

PP4 Considering only the completed orders, what is the average throughput time from the placement of the purchase order to a given stage of the process (either the receipt of the goods or the payment)?

PP5 What is the end-to-end performance of the P2P process (considering the entire chain of documents related to the order, which may or may not include purchase requisitions, invoices, and payments)?

PP6 Which are the activities that are correlated with a high processing time?

PP7 Does a high workload (in terms of open documents) in the purchasing or accounts payable departments lead to higher processing time?

The following goals aim to analyze *Process Compliance* (PC), where the goal is to make sure that actual executions are following the designed process model:

PC1 Does maverick buying (the order is placed without proper approval, and created in the system only after the invoice is received) happen in the process?

PC2 How many purchase requisitions were changed after the placement of the corresponding purchase order, just to match the information?

The last category is related to *Process Quality* (PQ), where the goal is to evaluate the quality of individual tasks in a process, plus identifying unintended behavior in the control flow:

PQ1 How many orders contain more than an invoice, leading to additional work in the accounts payable department?

PQ2 What are the unexpected patterns in the execution of the targeted P2P process?

The chosen goals were designed with a strong emphasis on addressing business-critical issues and enabling process optimization within the purchase-to-pay (P2P) process, in particular:

- *Alignment with Business Objectives*: All of the selected goals directly relate to crucial business objectives, such as improving efficiency, ensuring compliance, and enhancing the quality of operations. For instance, goals under

the Process Performance (PP) category aim to identify potential bottlenecks, inefficiencies, and areas for improvement in terms of capacity, throughput time, and workload management. These insights can be pivotal in shaping strategies for process optimization.

- *Focus on Compliance*: The Process Compliance (PC) goals were designed to address the risk of non-compliance to business rules and standards, which can result in costly mistakes, penalties, and reputation damage. By identifying instances of maverick buying and unauthorized modifications to purchase requisitions, businesses can enforce stricter controls and corrective measures to enhance compliance.

- *Quality Assurance*: Process Quality (PQ) goals focus on the quality and consistency of the P2P process. These insights are essential to maintaining high standards of operational quality and identifying patterns of errors or deviations that may affect overall performance.

The objectives of our analysis were initially formulated based on challenges and goals provided by our process users. To supplement those ideas, we conducted a sequence of seminars with experienced process analysts specifically designed to offer a preliminary understanding of process mining challenges pertinent to the purchase-to-pay (P2P) process. These seminars also introduced the foundational aspects of object-centric process mining. After these informational seminars, we initiated a brainstorming session. This interactive session highlighted the limitations and shortcomings of traditional process mining techniques when applied to the complex analysis of the P2P process. The valuable insights derived from this productive session played a crucial role and were instrumental in developing our comprehensive list of objectives. To narrow down our study and focus on some of the areas, we developed a small excerpt of questions for each of those areas.

The goals proposed here have been also carefully constructed to evaluate the efficacy of object-centric process mining within the context of the purchase-to-pay (P2P) process. They are designed to scrutinize key characteristics that make object-centric process mining a promising approach for gaining insights from process data, as opposed to traditional activity-centric process mining:

- Object-centric process mining allows the capture of intricate relationships among multiple data objects, thus enabling the study of complex business processes from multiple perspectives. The goals under Process Performance (PP) and Process Quality (PQ) categories are particularly aimed at this. For instance, understanding the processing capacity of the accounts payable department or examining the ratio of no-touch orders fundamentally involves inspecting how different business objects (like

invoices or orders) interact and evolve over time. This aligns well with the core strength of object-centric process mining.

- Object-centric process mining supports detailed conformance checking in a manner that is not restricted to sequences of activities. The goals under the Process Compliance (PC) category reflect this aspect. For example, identifying instances of maverick buying or changes in purchase requisitions after an order placement are all geared towards ensuring that business processes comply with preset rules and norms. Object-centric process mining enables this by not only considering the order of activities but also the state transitions of the related business objects.

We propose in Sect. 6 some analyses on the object-centric event log that allow fulfilling the aforementioned goals. The results are presented in Sect. 8.

4 Extraction

In this section, we describe our methodology for extracting object-centric event logs of purchase-to-pay (P2P) processes from Enterprise Resource Planning (ERP) systems, with a specific focus on SAP ERP systems. While our detailed extraction methodology is tailored to SAP ERP, it is important to note that the principles and strategies underpinning the extraction process can be generalized to other ERP systems as well. The specific SAP tables used for data extraction are reported here for replicability and transparency, but analogous tables or data structures can typically be found in other ERP systems.

The extraction process encompasses a range of activities, from identifying the different stages of the process, to extracting objects, relationships, and events. However, the extraction process in any ERP system is likely to encounter several challenges, such as scattered data across numerous tables, weak relational schemas, and multi-tenancy issues. Despite these challenges, our methodology can provide a reliable pathway to obtain object-centric event logs from a multitude of ERP systems, ultimately contributing to a broader applicability and understanding of object-centric process mining in business process management.

Each of the steps we illustrate in this section is based on the specific functionalities, characteristics, and data organization of SAP ERP systems. Nonetheless, these steps broadly align with what would be required in any ERP system for the extraction of an object-centric event log. Thus, while our examples may be SAP-specific, the process and its inherent logic can be extended to other contexts, serving as a blueprint for similar extraction efforts in different ERP environments.

In this work, we demonstrate the extraction of object-centric event logs from SAP ERP systems through a manual querying approach. The process involves detailed steps of identifying documents, objects, relationships, and events, as well as extracting these components from specific SAP tables. This manual process provides a high-quality, bespoke event log, and gives us a precise control over what data is extracted, from where, and how it is processed.

However, it is important to acknowledge the potential for automation in this process, a possibility explored in Berti et al [3]. Indeed, an approach for automatically extracting object-centric event logs from SAP ERP systems is proposed. While the quality of the resulting event logs may be slightly inferior to the ones obtained through the manual approach we demonstrate here, the automated process has distinct advantages in terms of efficiency, particularly for unknown processes or more frequent extraction tasks.

For the sake of simplicity, we limit the scope of our SAP analysis using the following criteria:

- A specific organization/company code.
- The control-flow perspective (activity-timestamp of the events) without any data attribute.
- The entire procurement stage of the process, and the invoicing-payment part of the accounting.
- A time interval.

We report the following system-specific challenges when extracting event data from SAP ERP:

- The relational schema of SAP ERP counts hundreds of thousands of tables. Even data related to popular processes (O2C, P2P) is scattered along many different tables.
- The relational schema of SAP ERP is weak, and foreign keys are usually implicitly defined and maintained at the application level.
- SAP ERP is a multi-tenant system and therefore events of different clients on the application side can be contained in the same tables. They can be distinguished by the client field (MANDT).
- The identifier of some documents is not unique but is unique in a given fiscal year. That poses the challenge that only the concatenation between the identifier and the fiscal year provides a unique reference to the document.
- The granularity of the timestamp: while for some information the timestamp is recorded with the second's granularity, for some other information the timestamp has a day's granularity.

To extract an object-centric event log for the considered P2P process, the following steps have been conducted:

Table 2 The different object types considered for the extraction of an object-centric event log for the P2P process

Object type description	Technical name in the log	Source table	Object identifier column(s)
Purchase requisition	Req.Doc	EBAN	BANFN
Purchase requisition item	Req.Item	EBKN	BANFN + BNFPO
Purchase order	Ord.Doc	EKKO	EBELN
Purchase order item	Ord.Item	EKPO	EBELN + EBELP
Invoice	Inv.Doc	RBKP	BELNR + GJAHR
Invoice Item	Inv.Item	RSEG	BELNR + BUZEI + GJAHR
Payment	Pay.Doc	BKPF	BELNR + GJAHR
Payment Item	Pay.Item	BSEG	BELNR + BUZEI + GJAHR
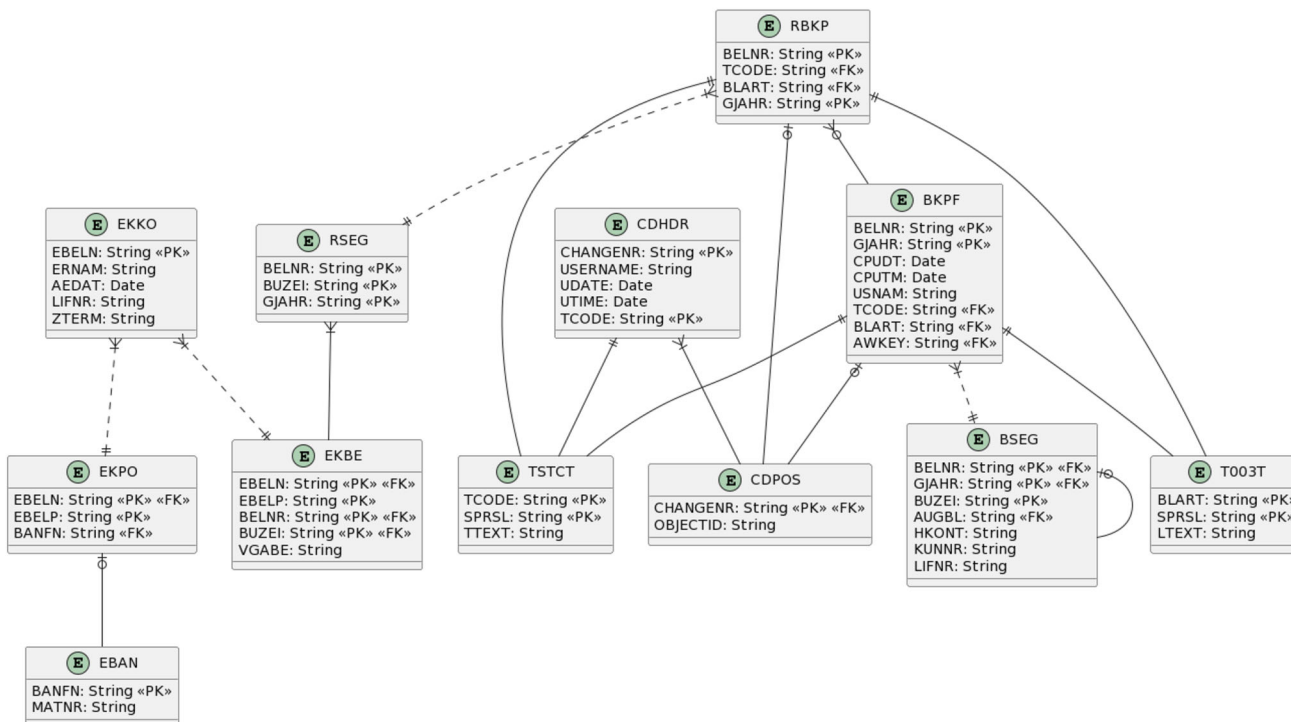
Fig. 5 Diagram showing the relationships between different tables related to the P2P process

1. *Identification of the different documents/stages of the process*: we identified the procurement stage (including the management of purchase requisitions and purchase orders documents, and the receipt of the goods) and the accounts payable stage (receipt and management of invoices and payments). The cardinalities of the relationships between these documents are summarized as follows:

   - Many-to-many relationships between purchase requisitions and purchase orders.
   - Many-to-many relationships between purchase orders and goods receipts.
   - Many-to-many relationships between purchase orders and invoices.

   - Many-to-many relationships between invoices and payments.

   The *xSuite* workflow system is responsible for the digital acquisition of documents related to purchase orders, goods receipts, and invoices. Since the correctness of the data inserted for the procurement stage is essential for the correct execution of the accounts payable process, the workflow system plays a big role in ensuring that data is inserted correctly in the system.

1. *Identification of the objects related to the given documents*: we consider the following correspondence:

- A purchase requisition document is associated with a "purchase requisition" object and some "purchase requisition item" objects.
- A purchase order document is associated with a "purchase order" object and some "purchase order item" objects. Moreover, if purchase requisitions are associated with the order, their "purchase requisition" and "purchase requisition item" objects are related to the order.
- A goods receipt document is associated with the corresponding "purchase order" and "purchase order item" objects (of all the received items).
- An invoice document is associated with a "invoice" object and some "invoice item" objects. Given the purchase orders associated with the invoice, their "purchase order" and "purchase order item" objects are associated with the invoice.
- A payment document is associated with a "payment" object and some "payment item" objects. Given the invoices associated with the payment, their "invoice" and "invoice item" objects are associated with the payment.

3. *Extraction of the objects*: after identifying the different object types, we identified the tables inside SAP ERP that can be used to extract the objects. The correspondence between the object types defined in the log and tables/objects are described in Table 2.

4. *Extraction of the relationships between the objects*: for this step, some tables of SAP ERP are used to link the different objects (see Table 3), as visualized in Fig. 5. All links are trivial except the invoice-payment connection.[2]

5. *Extraction of the basic events*: for this step, we extract some basic events directly from the entries of the considered tables as described in Table 4. We note that from the same table, events with different activities and timestamps can be extracted. For the analysis, we do not look for additional attributes aside from the activity and the timestamp. Since the required information is stored with the day's granularity (e.g., purchase requisition, purchase order, and invoice could happen on the same day), we add a time delta to the dates to ensure the correct logical order between the events.

6. *Extraction of the change events*: the rows of the table CDPOS become change events of our object-centric event log, reporting as related object identifiers the one cited in the table. The activity associated with these events

---

[2] The document number of the invoice reported in the accounting tables BKPF and BSEG is different from the number of the invoice in the RBKP table and can be linked based on the AWKEY field of the BKPF table.

Table 3 Link tables between the different object types

Link table	Source object type	Target object type
EKPO	Ord.Item	Req.Item
RSEG	Inv.Item	Ord.Item
BSEG	Pay.Item	Inv.Item**

These define relationships at the item level, which implicitly define relationships at the document level

depends on the fields that are created/updated/removed by the change.

5 Data preprocessing

We obtained an object-centric event log for the P2P process as described in Sect. 4. However, this object-centric event log might contain incomplete executions of the process (open orders still being processed in the system, or orders starting outside the considered time interval). Therefore, in order to get reliable insights for the analyses proposed in Sect. 6, we need to filter out incomplete executions of the process from the object-centric event log.

To do so, we adopt a simple data preprocessing strategy, which allows us to sample and filter the object-centric event log.

The strategy consists of building a graph (called the *object interaction graph*) in which the objects of the object-centric event log are the nodes, and the interactions between the objects (two objects belonging to the set of related objects of any event in the log) are the edges of this graph.

Starting from the object-centric event log contained in Table 1, we build the graph represented in Fig. 6. In this case, the different colors at the node level highlight different connected components of the given graph. The object-centric event log can be sampled keeping only the events related to the objects of a given connected component, or some properties of the connected components can be considered for filtering.

This allows us to easily filter out incomplete behavior (given that we applied a filter at the timestamp level). In our case, we would like to analyze the end-to-end process going from the purchase requisition to the payment. Since not all the orders are inserted with a purchase requisition, we want to consider all the connected components that contain at least a purchase order object and a payment object. In the example object-centric event log proposed in Table 1, we highlight the components that satisfy such criteria:

O = { { PR1 }, **{ PR2, PO1, R1, P1 }** , **{ PR3, PO2, R2, P2 }** , **{ PO3, R3, P3, R4, P4, R5, P5 }** , **{ PR4, PO4, GI1, R6, P6 }** , { PO5, R7 }, { PO6, PO7, R8 } }

Table 4 Basic activities extracted in the object-centric event log

Activity	Source table	Involved objects	Timestamp column	Time Delta
Create Purchase Requisition	EBAN	Req.Doc.:1 Req.Item:N	BADAT	+0.0s
Create Purchase Order	EKKO	Ord.Doc.:1 Ord.Item:N Req.Doc.:N Req.Item:N	AEDAT	+1.0s
Goods Receipt	EKBE	Ord.Doc.:N Ord.Item:N	BUDAT	+2.0s
Invoice Receipt	EKBE	Inv.Doc.:N Inv.Item:N Ord.Doc.:N Ord.Item:N	BUDAT	+3.0s
Invoice Created by Supplier	RBKP	Inv.Doc.:1 Inv.Item:N Ord.Doc.:N Ord.Item:N	BLDAT	+4.0s
Invoice Posted	RBKP	Inv.Doc.:1 Inv.Item:N Ord.Doc.:N Ord.Item:N	BUDAT	+5.0s
Clearance Posted (Payment)	BKPF	Pay.Doc.:1 Pay.Item:N Inv.Doc.:N Inv.Item:N	BUDAT	+6.0s

Fig. 6 Graph representing the interactions between the objects, starting from the object-centric event log contained in Table 1
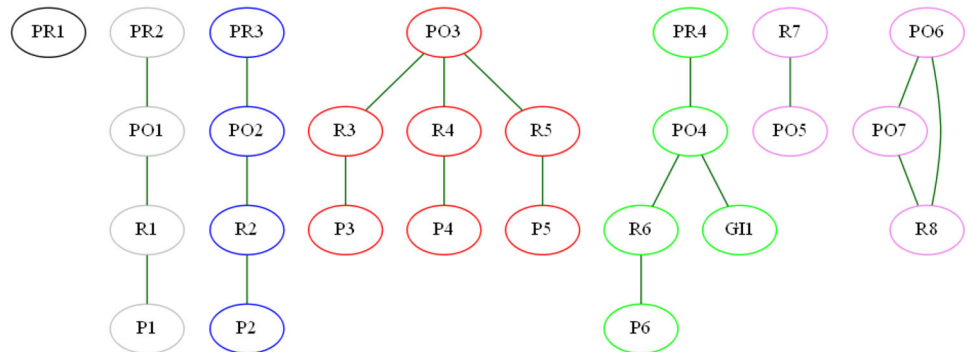
6 Mining and analysis

In this section, we describe the analyses which were adopted in the recent past to analyze the P2P process in the company (Sect. 6.1) and propose new object-centric analyses to address the analysis' goals (Sect. 6.2).

6.1 Previously adopted process discovery techniques

Existing techniques have been applied to different event log extractions, and we divide them between *traditional/case-centric techniques* (considering traditional event logs and the interaction between them) and *object-centric process mining techniques* (which require object-centric event logs as input).

Traditional/case-centric process mining: When working with traditional process mining techniques, a case notion needs to be chosen. The *order* or the *invoice* is popularly chosen as case notions for the P2P process. This results in event logs containing, for each case, all the events related to the processing of the order/invoice documents, plus all the events related to the processing of interconnected documents. For example, choosing *order* as a case notion would lead to including in the same case the events related to the purchase requisitions, goods receipts, invoices, and payments related to the order, in addition to the processing of the given order. This leads to "spaghetti" process models with unreliable annotations due to the convergence and divergence issues [22]. For example, Fig. 7 contains a "spaghetti" directly-

follows graph obtained from our P2P process. Simplification techniques can be applied to obtain simpler graphs; however, annotations are even more unreliable [21].

Commercial vendors are aware of the problems of traditional process mining techniques and tried to propose solutions. For example, Celonis supports the usage of "Multi-Event Logs" (MEL) in which interactions are shown between different process maps computed on different event logs of interconnected processes. (The interacting cases are assumed to be known.) These interactions are computed using temporal (*interleaved* & *non-interleaved miner*) and attribute-based criteria (*match miner*). Figure 8 contains an example multi-event log, in which the interleaved miner has been used to mine temporal interactions between the two process maps (the right one related to the "procurement" log, the left one related to the "accounts payable" log). We could see different interactions, including:

- An interaction between the *Create Purchase Requisition*, the *Create Purchase Order*, and the *Vendor Creates Invoice* activities, considering the order in which they are created. If goods are ordered prior to issuing of Purchase Requisition and Purchase Order, then *maverick buying*[3] happens. Problems connected to maverick buying can lead to a number of issues, such as overspending or lack of proper approvals. Tracking the temporal interactions between different process maps using the interleaved
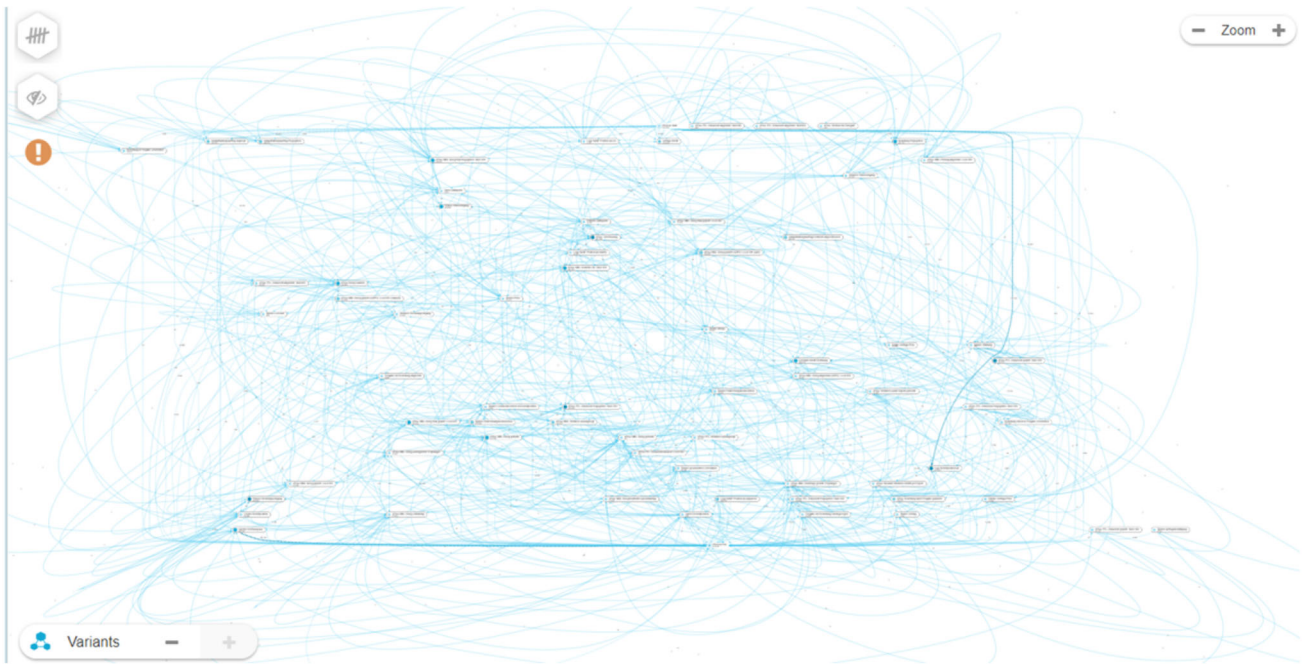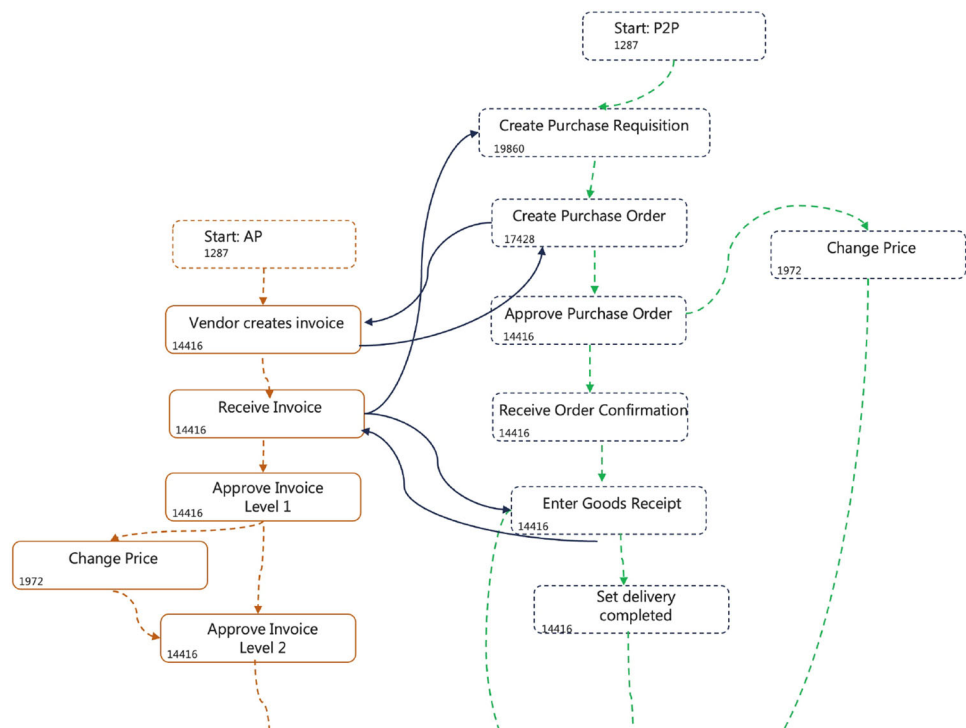
---

[3] https://easy-software.com/en/newsroom/glossary/maverick-buying/

Fig. 7 "Spaghetti" directly-follows graph of the P2P process obtained using traditional process mining techniques

Fig. 8 "Multi-event log" visualization showing the interaction between different logical components of the process (procurement and accounts payable)

miner can further improve visibility and understanding of the overall purchase process.

- An examination of the interactions between the *Enter Goods Receipt* and *Receive Invoice* activities is crucial to understand how the business users are processing goods receipts in practice. In theory, the goods receipt processing transaction (MIGO, in SAP) should be used to

confirm the quantity and quality of received materials. However, if it is observed that there is only a minimal time gap between the receipt of an invoice and entering the goods receipt activity, the users might be entering both invoice information and goods receipt information simultaneously. This interaction can be used to identify categories of goods for which the goods receipt activity is

unnecessary and only pro forma. This can help organizations to optimize their processes, eliminate unnecessary steps, and reduce costs. Additionally, it can also help to identify potential errors or discrepancies in the invoicing and goods receipt process and improve the accuracy of financial and inventory records.

However, while the meaning of some interactions discovered by the method was clear, some others are quite difficult to interpret.

Object-centric process mining: we considered mainly open-source prototypal software because commercial vendors (Celonis, MPM, IBM) are starting to offer support for object-centric process mining. For example, the Celonis vendor recently introduced the *Process Sphere* feature, which allows "to analyze and visualize the complex relationships between events and objects across interconnected processes"[4] and can ingest object-centric event logs in the OCEL standard. However, the features were still not available in our Celonis instance at the moment in which the analysis was performed.

Prototypal tools supporting the application of object-centric process mining techniques on top of OCELs are OCPM [4] and OCπ [1]. In Fig. 9, an object-centric directly-follows graph computed (using the OCPM tool) on an extract of the object-centric event log of the P2P process is shown. This contains different object types and activities (detailed in Tables 2 and 4). According to the provided abstraction, *Create Purchase Order* is preceded by *Create Purchase Requisition* and followed by *Invoice Created by Supplier*. Then, the goods and invoices are received by the company and recorded in the system (*Goods Receipt* and *Invoice Receipt*). The invoice is eventually posted for verification (*Invoice Posted*) and then the payment occurs clearing the different items of the invoices (*Clearance Posted (Payment)*).

An object-centric directly-follows graph (such as the one in Fig. 9) helps to identify the paths between the activities in a process. These paths can be annotated with frequency (as in the figure) or performance metrics. Therefore, we can get reliable statistics about the number of occurrences of an activity, or the number of objects related to events of the given activity. We could also identify the bottlenecks in the process by considering the average times between two directly following activities.

Limitations of existing object-centric techniques: in the context of a P2P process, the object-centric directly-follows graph allows us to visualize the performance and the compliance of the process:

- Maverick buying can be quantified by looking at the sum of the frequency of the input arcs of type *Ord.Doc.* in the *Create Purchase Order* activity.
- Post-mortem changes to purchase requisitions can be identified by looking at the output arcs of type *Req.Doc.* in the *Create Purchase Order* activity.

However, some analyses cannot be answered by considering object-centric directly-follows graphs. Regarding process performance, we are interested in the throughput time of a given step of the process (for example, the invoice processing time, which can be computed by the average difference between the completion and starting time for the objects of type invoice), or the end-to-end performance (from purchase requisitions to the completion of the payment) rather than the times of an individual path. Also, the correlations between the execution of an activity and the total throughput time, or the workload metrics (as the number of concurrently worked items) cannot be visualized on object-centric directly-follows graphs.

6.2 Novel analysis techniques

Mainstream object-centric process discovery techniques are not able to satisfy some of our goals; therefore, we focus on two novel analyses which allow us to retrieve the answers:

- *Graph-based analyses*: we used the implementation provided in the tool OCPM (https://www.ocpm.info/). Alternative implementations have been evaluated using PM4Py (https://pm4py.fit.fraunhofer.de/) and the Neo4J graph database [8].
- *Statistical analyses*: in this case, we used the tool OCPM to build a feature table on the object-centric event log and apply the available techniques of the tool to get some insights.

6.2.1 Graph-based analyses

To answer some of the analysis' goals, we can build an alternative directed graph (in comparison with the one introduced in Sect. 5) showing the short- and long-term dependencies between the objects in the object-centric event log. This is done by considering as nodes the objects of the log, and using the following criteria to connect objects with a directed arc:

- The objects should be both related to a given event.
- The source object's lifecycle does not start with the given event.
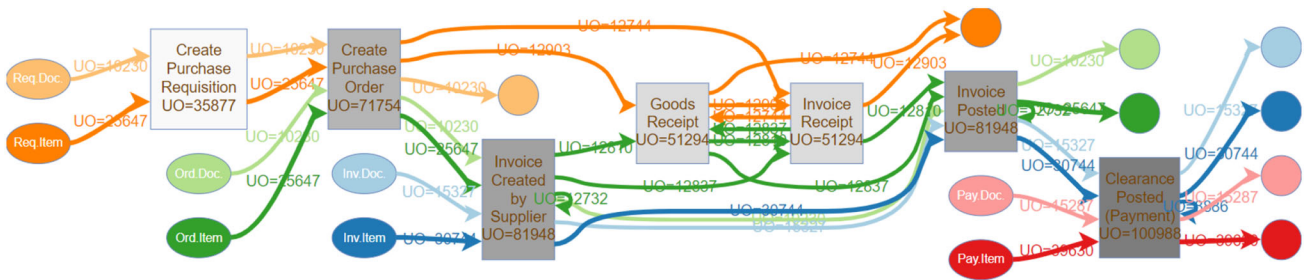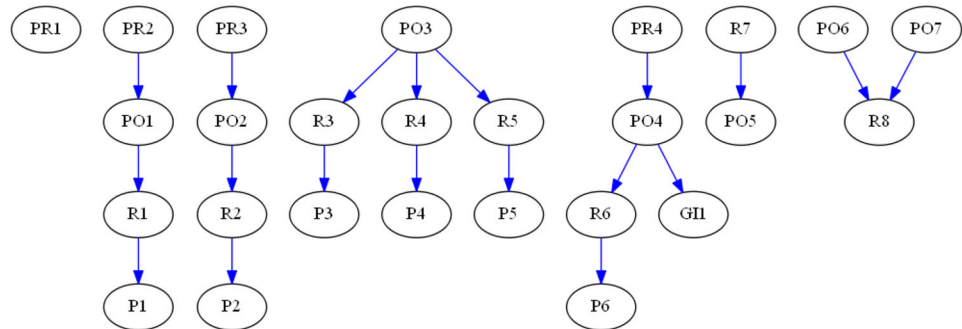- The target object's lifecycle starts with the given event.

Fig. 9 Object-centric directly-follows graph computed on an extract of the object-centric event log of the P2P process, annotated with the frequency of the arcs

Fig. 10 Example object creation graph computed starting from the object-centric event log contained in Table 1

We call this *object creation graph*. An example of an object creation graph computed starting from the object-centric event log contained in Table 1 is contained in Fig. 10.

The object interaction graph and the object creation graph serve distinct yet complementary purposes in the analysis of object-centric event logs, and their difference is pivotal to understand the flow and dependencies among objects.

The object interaction graph, first introduced in Sect. 5, plays a crucial role in discovering sets of interconnected objects. It is constructed based on events in which multiple objects participate together. It does not take into account the chronological order of the events, but merely the co-occurrence of objects within the same event. This graph is primarily useful for advanced filtering of data, where we might be interested in subsets of objects that interact directly with each other.

On the other hand, the object creation graph is designed to shed light on the logical progression of the objects in the object-centric event log. It helps to pinpoint the sequence in which objects are created and linked over time, based on the lifecycle of the objects. If an object is related to another and its lifecycle starts earlier, a directed edge from the first to the second object is established. This graph is instrumental in identifying long-term dependencies between objects and revealing a timeline of object creation and linkage.

Thus, while Fig. 6 (representing the object interaction graph) provides valuable information on object interactions based on event co-participation, it does not provide insights into the logical flow of object creation and linkage. Con-

versely, Fig. 10 (depicting the object creation graph) fills this gap by providing a chronological perspective on object relationships, allowing for the detection of long-term dependencies. Therefore, both graphs, despite their differences, collectively contribute to a more comprehensive understanding of object-centric event logs.

The goals that could be answered using a graph-based approach follow:

- For **PP4** and **PP5**, the object creation graph serves as a representation of long-term dependencies between different objects. Each node in this graph represents an object, and each edge represents a temporal relationship between the objects, such that the lifecycle of the source object is earlier than the lifecycle of the target object. By traversing this graph and studying the patterns and structures, we can gain insights into long-term dependencies that can inform the process performance analysis.
- **PC1** deals with identifying instances of maverick buying, which deviate from the standard order-to-invoice flow in the P2P process. Using the object creation graph, we identify these deviations by looking for edges where an invoice object is followed by an order object, which is the reverse of the expected order.
- In the case of **PC2**, the goal is to analyze the duration between the creation of purchase requisitions and their corresponding orders. This analysis is facilitated by the object creation graph, where a direct connection between a purchase requisition and an order indicates a valid pair

for this analysis. The timestamps of the first and last events in the lifecycles of these connected objects can then be used to compute the duration.

- For **PQ1**, we aim to identify orders that have been continued into more than one invoice, indicating possible quality issues. In the object creation graph, this scenario is represented by an order object node having outgoing edges to multiple invoice object nodes. By searching for such patterns in the graph, we can identify and analyze orders linked to multiple invoices.

6.2.2 Statistical analyses

In this section, we want to perform specific analyses (which could not be performed by interpreting the object-centric directly-follows graph) using statistical analyses offered by the *Correlation Statistics* (first two analyses) and *Conformance* pages of the machine learning component of the OCPM tool:

- **PP6** seeks to determine the activities that are correlated with a high processing time. This analysis is performed by using correlation statistics, specifically the correlation coefficient or Goodman–Kruskal's gamma measure, which are suitable for assessing associations in one-hot encoded categorical variables. For each activity, a binary feature is created indicating its occurrence, and the correlation between these binary features and the processing time (target variable) is computed to identify those activities that are associated with high processing time.
- In addressing **PP7**, the question is whether a high workload, as represented by the number of open documents, leads to higher processing time. Here, correlation measures such as Kendall's Tau or Spearman's Rho are used, as these nonparametric measures are more suitable for analyzing the monotonic relationship between two continuous variables. A continuous feature representing the number of concurrently worked objects (workload) is computed, and its correlation with processing time (target variable) is examined to determine whether a high workload indeed leads to higher processing times.
- The objective of **PQ2** is to identify outlier patterns in the execution of the business process. Anomaly detection techniques, specifically isolation forests, are used to distinguish between anomalous and non-anomalous objects. The output of the isolation forest is an anomaly score for each object. The objects are then labeled as anomalous or non-anomalous based on a chosen threshold for the anomaly scores. Classification rules are subsequently discovered from these labels using rule-based machine learning techniques, which provide a transparent way to characterize the patterns associated with anomalies.

Table 5 SQL queries executed in the OCPM tool to respond to some analysis' goals

Goal	SQL Query
PP1	SELECT Count(*) FROM OCEL_OBJECTS WHERE obj_lif_act_Invoice_Receipt=1 AND object_interaction_ot_InvDoc=0
PP2	SELECT AVG(obj_lif_dur) FROM OCEL_OBJECTS WHERE obj_lif_act_Invoice_Receipt=1 AND object_interaction_ot_InvDoc=0
PP3	SELECT Count(*) FROM OCEL_OBJECTS WHERE obj_lif_length=1 AND obj_lif_act_Create_Purchase_Order=1 AND object_interaction_ot_OrdDoc=0

This section highlights the essential role of statistical analyses when used in combination with the object-centric process mining approach, addressing analysis' goals that are beyond the interpretive capacity of the object-centric directly-follows graph alone. The application of correlation measures to assess the relationship between activities and processing time, as well as between workload and processing time, provides insights into performance dynamics within the process. In addition, employing anomaly detection methods helps uncover uncommon patterns in process execution, thus enabling a comprehensive characterization of the process. These analyses contribute to an enriched understanding of the process, crucial for in-depth process analysis and further research endeavors in object-centric process mining.

Table 5 provides a set of SQL queries that were executed in the OCPM tool (*Machine Learning - SQL Explorer*) in order to address several specific analysis goals related to our study. The queries correspond to the following performance and conformance analysis goals:

- The goal **PP1** aimed to evaluate the current processing capacity of the accounts payable department. We addressed this by counting the number of objects that went through the 'Invoice Receipt' activity but did not interact with 'InvDoc' (so they are *Inv.Doc.* themselves).
- The second goal, **PP2**, targeted the evaluation of the time required by the accounts payable department to process and verify a single invoice document. For this, we calculated the average lifecycle duration of the same set of objects considered in **PP1**.
- Finally, the third goal, **PP3**, sought to measure the number of procurement orders that were correctly inserted at the creation phase, negating the need for further changes (termed 'no-touch' orders). This was assessed by counting the number of objects that underwent only the 'Create Purchase Order' activity and had no interaction with 'OrdDoc' (so they are *Ord.Doc.* themselves).

In each case, the queries facilitated an effective and tailored analysis of specific performance and conformance objectives in the P2P process.

7 The OCPM tool

In this section, we present the main features of the OCPM tool that was used to perform the analysis. The Javascript-based tool is publicly available at https://www.ocpm.info/ and a demo on a simulated log is available at https://www.ocpm.info/ocel_demo.html.

The OCPM tool provides a rich set of object-centric process mining features:

- Ingestion/exporting of object-centric event logs in the OCEL standard format (JSON-OCEL and XML-OCEL).

- Flattening the object-centric event logs into traditional event logs with the choice of a case notion.
- Advanced preprocessing features (filtering, sampling).
- Discovering object-centric process models: object-centric directly-follows graphs [4], and object-centric [19].
- Conformance checking on object-centric event logs based on declarative and temporal constraints (log skeleton, temporal profile).
- Exploration of the events/objects of the object-centric event log.
- Machine learning (anomaly detection, correlation analytics, advanced conformance checking).

Figure 11 shows the process model component of the proposed tool on top of the "demo" dataset. The position of different logical components is highlighted. In particular, (1) allows for the selection of the object-centric process model

Fig. 11 Overall view over the process model page of the OCPM tool
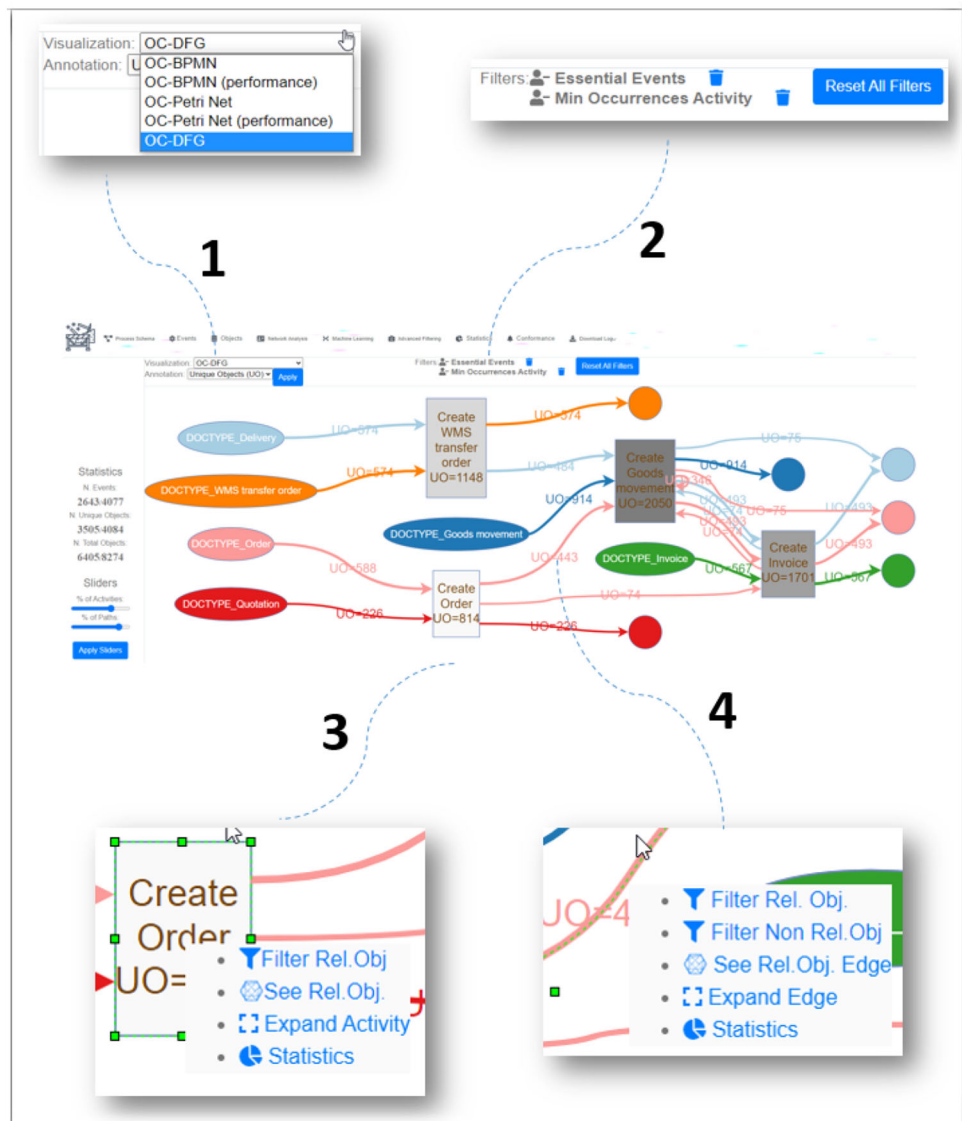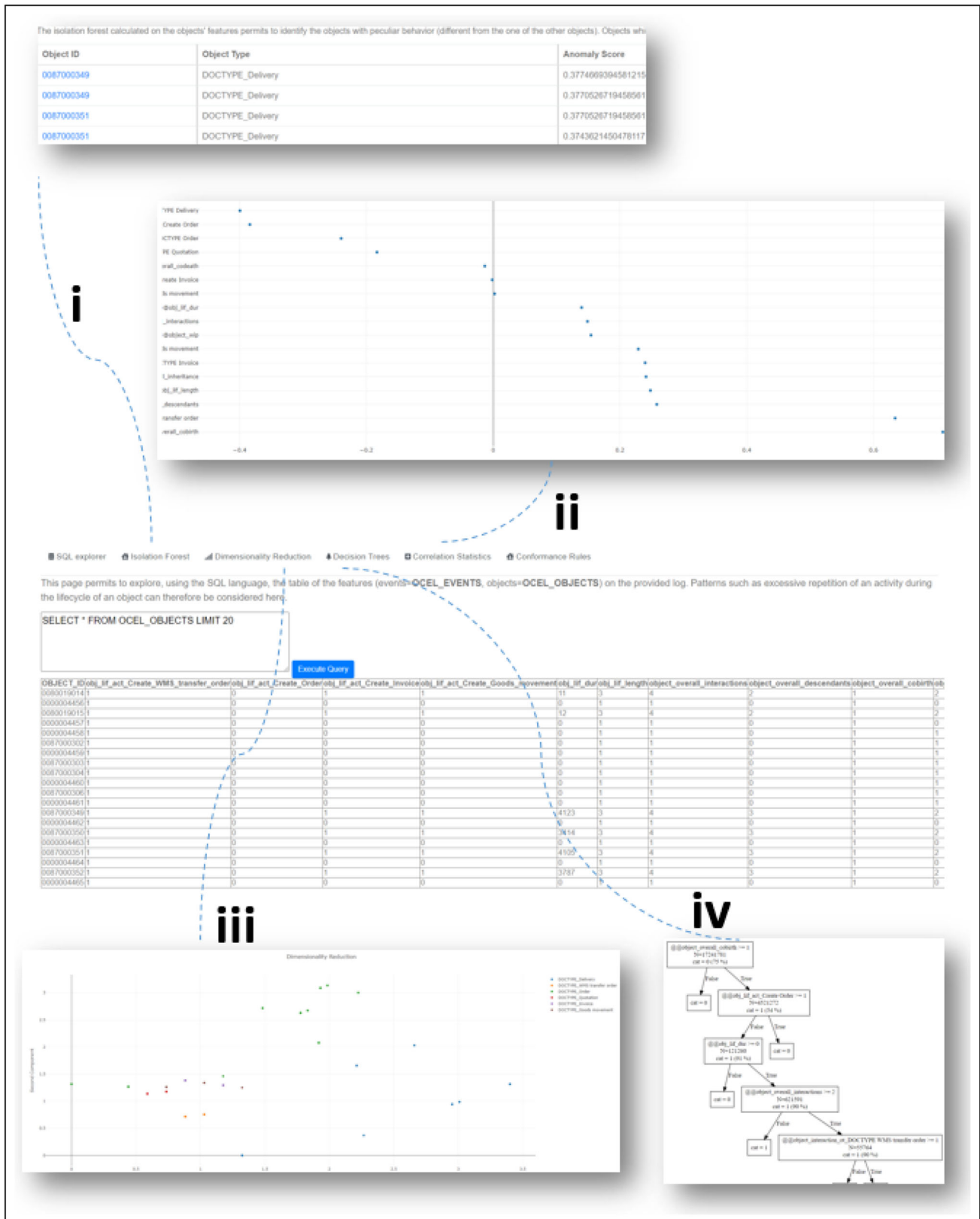
Fig. 12 Machine learning component of the OCPM tool

to discover, (2) allows to see and remove the filters applied on the event log, (3) allows to apply a filter on an activity of the model, (4) allows to apply a filter on an edge of the model.

Figure 12 shows the machine learning component of the proposed tool on top of the "demo" dataset. The landing page computes the features of the object-centric event logs and proposes an SQL explorer to navigate/query the values for the different objects. In (i), an anomaly detection algorithm is applied to the features in order to identify the anomalous objects (click an object allows to explore its lifecycle). In (ii), correlation statistics are computed between a feature and the other features of the object-centric event log. In (iii), a dimensionality reduction technique (FASTMAP) is used to show groups of objects with similar features. In (iv), the correlation between a feature and the other features is explored by means of decision trees.

8 Evaluation

In this section, we evaluate the correctness and the quality of the insights obtained by our analysis. These are the steps of our evaluation:

- Since we developed a custom extractor of object-centric event logs, we checked if the information extracted from SAP (number of objects, timestamps, the flow of documents) matches the result of some reporting transactions in SAP (in particular, *ME53N* (Display Purchase Requisition), *ME23N* (Display Purchase Order), and *FB03* (Display Finance Document)).
- We compared the frequency/performance annotations with the ones obtained by applying traditional techniques offered by the Celonis tool.

To make the comparisons feasible, we focused on smaller samples of the original object-centric event logs (obtained by applying the technique described in Sect. 5).

The analytical comparison shows some advantages of object-centric process mining over traditional process mining techniques:

- We could obtain the correct number of documents for every stage of the process (**PP1**). This applies in particular to the number of distinct purchase requisitions. Some purchase requisitions involved items that were then procured by different purchase orders. Therefore, the number of purchase requisitions obtained in the Celonis software was significantly bigger than the true number of purchase requisitions (due to the convergence problem).
- We found that the end-to-end performance (**PP5**) of the purchase-to-pay process was skewed by the reversal of some payments, which was not supported by the traditional extractor (because it would require other computationally expensive table joins), but that we could support in the object-centric setting.

Moreover, we found some interesting compliance/quality patterns in the considered timespan:

- Maverick Buying (**PC1**) is a deleterious behavior where the traditional approval steps of a purchase order are skipped and the order is placed directly with the supplier. An invoice is then received by the supplier, which leads to the creation of the purchase order. We could detect a non-negligible amount of maverick buying in our process.
- Post-Mortem Changes to Purchase Requisitions (**PC2**): another deleterious behavior observed in the process is that purchase requisitions are changed after their approval to match the amounts/quantities of the purchase order.

Table 6 Effect of some activities on the average end-to-end performance of the process

Activity	Occ.	End-to-End Perf. with Act.	General End-to-End Perf.	Increase
Change Receipt Indicator	5	2Y	6MO	3.1
Change Purchasing Group	10	2Y	6MO	3.9
Change Quantity	9	1Y	6MO	3.6
Change Commitments	3	1Y	6MO	3.6
Change Net Value	17	1Y	6MO	3.1
Change Gross Value	17	1Y	6MO	3.1
Change Effective Value	17	1Y	6MO	3.1
Assign PO Account	15	1Y	6MO	3.1
Tax code information	15	1Y	6MO	3.0
Change Net Price	13	1Y	6MO	2.9

Change activities involving orders/invoices contribute significantly to end-to-end performance

- Orders with Duplicated Invoices (**PQ1**): we could detect some orders with duplicated invoices in our system. Also, some orders with dozens of different invoices exist. After investigation, we detected that they are *maintenance contracts*. An example of a maintenance contract is a cleaning contract, repeated weekly at the same conditions and invoiced monthly.

In addition, also the statistical analyses provided us with interesting insights:

- We could detect a high correlation (**PP6**) between the presence of change activities (for both orders and invoices) with the processing time. In particular, changes in the amount (to pay) for the invoices, or the amounts expected for the orders, are particularly deleterious on the processing time (see Table 6).
- In our process and in the considered timespan, the workload is highly correlated (correlation coefficient 0.89) with the processing time (**PP7**). Therefore, an high work-
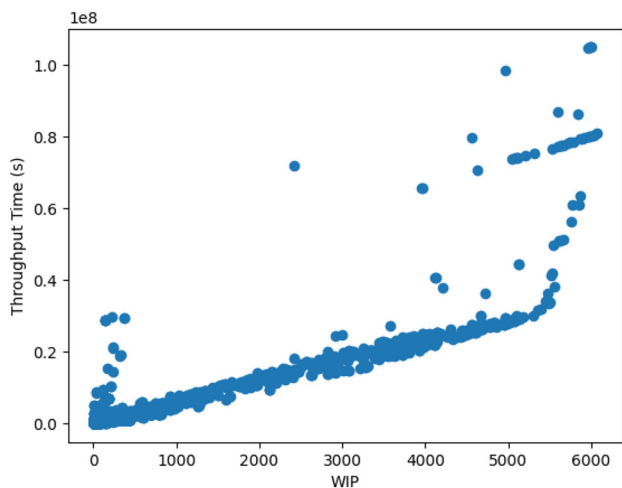

Fig. 13 Scatter plot showing the high correlation (0.89) between the workload (on the x-axis) and the throughput time (on the y-axis) of the objects of the event log

load in the process (number of open documents) leads to a higher processing time (see Fig. 13).
- Considering **PQ2**, e could identify some purchase orders with a significant number of change activities ($\geq$ 150) executed on the order. Moreover, we can identify invoices related to more than 20 orders.

In our evaluation, we also accounted for the extraction times for the diverse object types essential for the P2P process. As Table 7 demonstrates, these times varied, reflecting the inherent differences in complexity among the various SAP tables. The extraction time for purchase requisitions was approximately 10 s, while for more complex objects such as invoice items, the process extended to 95 s. These extraction times underline the computational requirements of the extraction process, critical to constructing a comprehensive view of the P2P process. It is important to note that these extraction times are influenced by numerous factors, such as system resources, data volume, and database design. Consequently, these times should be seen as indicative, and the exact figures may vary based on the specifics of the system and the data involved.

9 Improvements

The results of the analysis are currently being used to improve the execution of the business process. In particular, the following improvements have been already adopted in the company:

- Our study delved into the issues related to maverick buying (**PC1**) and postmortem changes in purchase requisitions (**PC2**), and through this investigation, we were able to identify steps to address these issues. These included updating internal documents with additional instructions and incorporating adjustments to internal training programs to improve the purchasing process.

Table 7 Extraction times for the different object types considered for the P2P process

Object type description	Technical name in the log	Source table	Object identifier column(s)	Extraction Time (sec)
Purchase requisition	Req.Doc	EBAN	BANFN	10
Purchase requisition item	Req.Item	EBKN	BANFN + BNFPO	15
Purchase order	Ord.Doc	EKKO	EBELN	30
Purchase order item	Ord.Item	EKPO	EBELN + EBELP	50
Invoice	Inv.Doc	RBKP	BELNR + GJAHR	40
Invoice Item	Inv.Item	RSEG	BELNR + BUZEI + GJAHR	95
Payment	Pay.Doc	BKPF	BELNR + GJAHR	200
Payment Item	Pay.Item	BSEG	BELNR + BUZEI + GJAHR	350

- By identifying activities that were correlated with prolonged processing times (**PP6**), we were able to conduct a series of workshops with business users and apply the Pareto principle to identify the main underlying causes. This allowed us to target the most significant contributors to the processing time, and work towards implementing effective solutions.

- Our analysis identified instances of duplicated invoices (**PQ1**), which provided insight into issues with master data management within the company. Further investigation revealed that these duplicates were often the result of workarounds implemented by individual users to speed up the invoice entry process. By addressing these root causes, the company was able to improve its master data management and reduce the occurrence of duplicated invoices.

- By examining the correlation between high workload and prolonged processing times (**PP7**), we were able to identify "peak times" when the workload was highest. With this information, we were able to find temporary resource solutions to stabilize the workload and ensure a more efficient process.

Other considerations are currently evaluated but still not enacted:

- By analyzing the results of **PP1** and **PP2**, we were able to discern distinct groups of invoices that exhibited particularly long or short lifecycle times. While some of these cases may be justifiable due to factors such as long-term contracts or emergency maintenance orders, we also identified outliers that cannot be explained by these criteria. From the perspective of business users, these outliers may be considered excessive. To address this issue, potential interventions may include updating relevant master data and conducting interdepartmental workshops to address underlying issues.

- By considering the results of **PP4** and **PP5**, we were able to determine whether invoices with prolonged lifecycle times were affected by issues related to purchase orders, such as long processing times or bottlenecks. This information was used to inform the design of improved communication strategies between the departments responsible for invoice payments and ordering goods, to address these identified issues and anticipate them in the future.

- The identification and examination of outliers in interconnected processes (**PQ2**) helped to initiate a discussion about the desired behavior in these processes, and what constitutes non-compliant behavior. This ultimately led to a deeper understanding of the processes and how to improve them.

- We have observed that the measure of 'no-touch' orders (**PP3**) could indicate efficiency within the procurement process. These are orders that were correctly entered during the creation phase, thus negating the need for further modifications. Although this factor has not been explicitly addressed in our improvement actions, it offers significant potential for enhancing process efficiency. An increase in the number of 'no-touch' orders signifies a reduction in time and resources spent on order modifications, leading to a more streamlined process. Therefore, we are considering strategies to improve the accuracy at the order creation stage, such as better training in order entry procedures and improved documentation. In addition, we are looking into technological solutions that could assist in avoiding errors during order creation, such as AI-based assistance or predictive algorithms.

10 Related work

The paper described process mining analyses on top of a purchase-to-pay process supported by SAP ERP. Therefore, we consider scientific results related to the extraction of traditional/object-centric event data and the results of previous case studies. Moreover, since the interconnected nature of objects in SAP, we consider graph-based techniques.

Extraction of traditional event logs from SAP ERP: SAP is an interesting system for process mining since its widespread usage by companies and the unstructuredness of the supported processes. Hence, several process mining publications targeted the extraction of data from SAP ERP. In Ingvaldsen and Gulla [13], a method is proposed for the extraction and transformation of event logs from SAP ERP, which involves the manual specification of a meta-model defining how events, resources, and their relationships are stored. The method has been applied to SAP systems provided by Norwegian Agricultural and Marketing Cooperative and Nidar. Some limitations exist: although all the information (transactional, master, and ontological data) needed to extract meaningful process models is available, the transactions are not mapped directly to the tasks. Moreover, it was not possible to map the extracted transaction flow to the processes in the SAP reference model. In de Murillas et al. [5], a meta-model that can ingest the contents of a relational database and provide the possibility to easily specify queries to produce an event log is described. The meta-model can be used on a database supporting SAP ERP. However, this leads to the generation of traditional event logs having convergence/divergence issues [22].

Extraction of object-centric event data from SAP ERP: Some approaches have been proposed to avoid the drawbacks of using traditional event logs. In Lu et al. [14], the construction of artifact-centric models on top of SAP ERP

is proposed, along with an implementation in the popular ProM 6.x framework. An artifact-centric model considers both the lifecycle of an artifact (purchase order document, invoice document, payment document) and the interaction between different artifacts. Some limitations exist: the approach requires some non-trivial manual steps, and the discovery phase is limited to two artifacts. In Berti et al. [3], a method to extract object-centric event logs starting from SAP ERP is proposed, which is the foundation of the current paper. The proposed prototypal software is limited by an in-memory approach and customization options. Moreover, some fundamental details (the construction of the GoR and the extraction of the relationships between the table entries) have been omitted for space reasons.

Process mining case studies on top of SAP ERP: SAP ERP stores interesting but company-critical data. Therefore, few case studies in applying process mining on top of SAP ERP data have been proposed. In Fleig et al. [11], an application of process mining to the Order-to-Cash and Procure-to-Pay processes of a manufacturing company is proposed. An application of process mining to the Procure-to-Pay and Accounts Payable processes is proposed in Stolfa et al. [17]; Stephan et al. [16]. The warehouse management process is considered in ER et al. [6]. Also, Fleig et al. [10] discusses the implementation of a decision support system, supported by process mining, for the standardization of ERP systems.

Graph-based analyses of SAP ERP: The graph-based nature of event data is exploited in Esser and Fahland [7]. Traditional (and object-centric) event logs can be encoded in a graph database. This allows for queries that are unfeasible on top of relational databases since edges are first-class entities in graph databases. An application to ERP systems (BPI Challenge 2019 log) is proposed. The contribution in Fahland [9] further exploits the graph- and object-based nature of event data to build event knowledge graphs. This data structure allows us to naturally model behavior over multiple entities as a network of events.

11 Conclusion

In this paper, we presented a case study concerning the application of object-centric process mining techniques to a real-life P2P process, along with novel ad hoc techniques that were needed to process the event data. We adapted the PM^2 process mining project methodology to the object-centric setting, in particular in the extraction (retrieval of the events, objects, and event-to-object relationships), pre-processing (proposing a simple filtering/sampling technique in the object-centric setting) and analysis (with the usage of object-process mining techniques). During the analysis, we considered both traditional and novel techniques. In particular, we proposed process-tailored graph- and statistical-based

paradigms. The results that we obtained were interesting with regard to the performance and compliance of the process (Sect. 8). Currently, we have enacted some of the insights, and some other optimizations are planned (Sect. 9).

The experience of applying object-centric process mining in a real-world purchase-to-pay (P2P) scenario provided us with a wealth of lessons learned. The extraction of an object-centric event log from the SAP ERP system proved to be challenging. Despite the guidance from Berti et al. [3], we faced issues with the complexity and the lack of standardization in the system's data structures. A considerable amount of time was invested in identifying the relevant tables and mapping their relationships to obtain a meaningful representation of the process.

In addition to that, defining object lifecycles and their interactions in the context of the P2P process required an intimate understanding of the domain. While the object interaction graph helped us discover sets of connected objects, the object creation graph was instrumental in identifying the logical flow of objects and long-term dependencies. The differentiation between the two and their respective uses was a significant learning point in our study.

Also, the execution of statistical analyses posed a challenge due to the inherent complexity of the data. Correlation analysis between different types of variables (continuous and categorical) required careful choice and application of statistical measures. Anomaly detection and the consequent extraction of classification rules were a delicate task, and while we were able to find some meaningful anomalies, we suspect that the implementation in the OCPM tool could benefit from more advanced anomaly detection techniques (for example, the Local Outlier Factor (LOF) method should theoretically work better).

The interpretation and explanation of the results to the domain experts and stakeholders were met with hurdles due to the technical nature of our findings. Our endeavor was to make the results as intelligible as possible for the non-technical audience, a challenge we tried to overcome with visual aids and simplifying the complex statistical results.

These challenges, while at times cumbersome, provided valuable insights into the real-world application of object-centric process mining, enlightening our future research and application in this domain.

Author Contributions UJ proposed the case study, performed the data extraction and analysis, and implemented changes to the real-life P2P process. AB provided support in the application of object-centric process mining to the P2P process, and wrote Section 2, 3, and 4, and contributed to Section 5. GP designed the organization of the case study and wrote Section 1. MR contributed to Section 5. Moreover, GP, MR and WA ensured the quality of the other sections of the paper and provided detailed feedback.

Data Availability Not applicable.

Declarations

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

Ethical approval Not applicable.

References

1. Adams, J.N., van der Aalst, W.M.P.: Ocπ: object-centric process insights. In: Bernardinello, L., Petrucci, L. (eds.) Application and Theory of Petri Nets and Concurrency—43rd International Conference, PETRI NETS 2022, Bergen, Norway, June 19–24, 2022, Proceedings, Lecture Notes in Computer Science, vol. 13288, pp 139–150. Springer, New York. https://doi.org/10.1007/978-3-031-06653-5_8 (2022)

2. Badakhshan, P., Bernhart, G., Geyer-Klingeberg, J., et al.: The action engine-turning process insights into action. ICPM Demos (2019)

3. Berti, A., Park, G., Rafiei, M., et al.: An event data extraction approach from SAP ERP for process mining. In: Munoz-Gama, J., Lu, X. (eds.) Process Mining Workshops—ICPM 2021 International Workshops, Eindhoven, The Netherlands, October 31–November 4, 2021, Revised Selected Papers, Lecture Notes in Business Information Processing, vol. 433, pp. 255–267. Springer, New York. https://doi.org/10.1007/978-3-030-98581-3_19 (2021)

4. Berti, A., van der Aalst, W.M.P.: OC-PM: analyzing object-centric event logs and process models. CoRR (2022). arXiv:2209.09725

5. de Murillas, E.G.L., Reijers, H.A., van der Aalst, W.M.P.: Connecting databases with process mining: a meta model and toolset. Softw. Syst. Model. **18**(2), 1209–1247 (2019). https://doi.org/10.1007/s10270-018-0664-7

6. ER, M., Astuti, H.M., Wardhani, I.R.K.: Material movement analysis for warehouse business process improvement with process mining: a case study. In: Bae, J., Suriadi, S., Wen, L. (eds.) Asia Pacific Business Process Management—Third Asia Pacific Conference, AP-BPM 2015, Busan, South Korea, June 24–26, 2015, Proceedings, Lecture Notes in Business Information Processing, vol. 219, pp. 115–127. Springer, New York. https://doi.org/10.1007/978-3-319-19509-4_9 (2015)

7. Esser, S., Fahland, D.: Multi-dimensional event data in graph databases. J Data Semant **10**(1–2), 109–141 (2021). https://doi.org/10.1007/s13740-021-00122-1

8. Fahland, D.: Multi-dimensional process analysis. In: Ciccio, C.D., Dijkman, R.M., del-Río-Ortega, A., et al. (eds.) Business Process Management—20th International Conference, BPM 2022, Münster, Germany, September 11–16, 2022, Proceedings, Lecture Notes in Computer Science, vol. 13420, pp. 27–33. Springer, New York. https://doi.org/10.1007/978-3-031-16103-2_3 (2022a)

9. Fahland, D.: Process mining over multiple behavioral dimensions with event knowledge graphs. In: van der Aalst, W.M.P., Carmona, J. (eds.) Process Mining Handbook. Lecture Notes in Business Information Processing, vol. 448, pp. 274–319. Springer, New York (2022)

10. Fleig, C., Augenstein, D., Maedche, A.: Designing a process mining-enabled decision support system for business process standardization in ERP implementation projects. In: Weske, M., Montali, M., Weber, I., et al. (eds.) Business Process Management Forum—BPM Forum 2018, Sydney, NSW, Australia, September 9–14, 2018, Proceedings, Lecture Notes in Business Information Processing, vol. 329, pp 228–244. Springer, New York. https://doi.org/10.1007/978-3-319-98651-7_14 (2018a)

11. Fleig, C., Augenstein, D., Maedche, A.: Process mining for business process standardization in ERP implementation projects—an SAP S/4 HANA case study from manufacturing. In: van der Aalst, W.M.P., Casati, F., Conforti, R., et al. (eds.) Proceedings of the Dissertation Award, Demonstration, and Industrial Track at BPM 2018 co-located with 16th International Conference on Business Process Management (BPM 2018), Sydney, Australia, September 9–14, 2018, CEUR Workshop Proceedings, vol. 2196, pp 149–155. CEUR-WS.org, Aachen. http://ceur-ws.org/Vol-2196/BPM_2018_paper_31.pdf (2018b)

12. Ghahfarokhi, A.F., Park, G., Berti, A., et al.: OCEL: a standard for object-centric event logs. In: Bellatreche, L., Dumas, M., Karras, P., et al. (eds.) New Trends in Database and Information Systems—ADBIS 2021 Short Papers, Doctoral Consortium and Workshops: DOING, SIMPDA, MADEISD, MegaData, CAoNS, Tartu, Estonia, August 24–26, 2021, Proceedings, Communications in Computer and Information Science, vol. 1450, pp. 169–175. Springer, New York. https://doi.org/10.1007/978-3-030-85082-1_16 (2021)

13. Ingvaldsen, J.E., Gulla, J.A.: Preprocessing support for large scale process mining of SAP transactions. In: ter Hofstede, A.H.M., Benatallah, B., Paik, H. (eds.) Business Process Management Workshops, BPM 2007 International Workshops, BPI, BPD, CBP, ProHealth, RefMod, semantics4ws, Brisbane, Australia, September 24, 2007, Revised Selected Papers, Lecture Notes in Computer Science, vol. 4928, pp. 30–41. Springer, New York. https://doi.org/10.1007/978-3-540-78238-4_5 (2007)

14. Lu, X., Nagelkerke, M., van de Wiel, D., et al.: Discovering interacting artifacts from ERP systems. IEEE Trans. Serv. Comput. **8**(6), 861–873 (2015). https://doi.org/10.1109/TSC.2015.2474358

15. Reinkemeyer, L. (ed.): Process Mining in Action: Principles, Use Cases and Outlook. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-40172-6

16. Stephan, S., Lahann, J., Fettke, P.: A case study on the application of process mining in combination with journal entry tests for financial auditing. In: 54th Hawaii International Conference on System Sciences, HICSS 2021, Kauai, Hawaii, USA, January 5, 2021. ScholarSpace, Denver, pp. 1–10. https://hdl.handle.net/10125/71314 (2021)

17. Stolfa, J., Kopka, M., Stolfa, S., et al.: An application of process mining to invoice verification process in SAP. In: Abraham, A., Krömer, P., Snásel, V. (eds) Innovations in Bio-inspired Computing and Applications—Proceedings of the 4th International Conference on Innovations in Bio-Inspired Computing and Applications,

IBICA 2013, August 22–24, 2013, Ostrava, Czech Republic, Advances in Intelligent Systems and Computing, vol. 237, pp. 61–74. Springer, New York. https://doi.org/10.1007/978-3-319-01781-5_6 (2013)

18. van der Aalst, W.M.P.: Process Mining: Data Science in Action, 2nd edn. Springer, New York (2016)

19. van der Aalst, W.M.P., Berti, A.: Discovering object-centric petri nets. Fundam Informaticae **175**(1–4), 1–40 (2020). https://doi.org/10.3233/FI-2020-1946

20. van Eck, M.L., Lu, X., Leemans, S.J.J., et al.: PM^2: a process mining project methodology. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds.) Advanced Information Systems Engineering—27th International Conference, CAiSE 2015, Stockholm, Sweden, June 8–12, 2015, Proceedings, Lecture Notes in Computer Science, vol. 9097, pp. 297–313. Springer, New York. https://doi.org/10.1007/978-3-319-19069-3_19 (2015)

21. van der Aalst, W.M.P.: A practitioner's guide to process mining: limitations of the directly-follows graph. In: Cruz-Cunha, M.M., Martinho, R., Rijo, R., et al. (eds.) CENTERIS 2019—International Conference on ENTERprise Information Systems/ProjMAN 2019—International Conference on Project MANagement/HCist 2019—International Conference on Health and Social Care Information Systems and Technologies 2019, Sousse, Tunisia, Procedia Computer Science, vol. 164, pp 321–328. Elsevier, Amsterdam. https://doi.org/10.1016/j.procs.2019.12.189 (2019b)

22. van der Aalst, W.M.P.: Object-centric process mining: dealing with divergence and convergence in event data. In: Ölveczky, P.C., Salaün, G. (eds.) Software Engineering and Formal Methods - 17th International Conference, SEFM 2019, Oslo, Norway, September 18–20, 2019, Proceedings, Lecture Notes in Computer Science, vol. 11724, pp. 3–25. Springer, New York. https://doi.org/10.1007/978-3-030-30446-1_1 (2019a)

23. vom Brocke, J., Mendling, J.: Frameworks for business process management: a taxonomy for business process management cases. In: vom Brocke, J., Mendling, J. (eds.) Business Process Management Cases, Digital Innovation and Business Transformation in Practice. Management for Professionals, pp. 1–17. Springer, New York. https://doi.org/10.1007/978-3-319-58307-5_1 (2018)