PROCESS MINING OF EVENT LOGS IN INTERNAL AUDITING: A CASE STUDY

Mieke Jans
Hasselt University, Belgium
Mieke.jans@uhasselt.be

Michael Alles
Rutgers Business School
Newark, NJ, USA
alles@business.rutgers.edu

Miklos Vasarhelyi
Rutgers Business School
Newark, NJ, USA
miklosv@andromeda.rutgers.edu

April 1, 2011*

Abstract Jans et al. (2010) argued that process mining has the potential to increase the effectiveness and scope of internal auditing by enabling auditors to extract knowledge from event logs recorded by a business's information system. In this paper we explore the value added that the techniques of process mining can provide to internal auditors by conducting a case study of the procurement process at a major European financial services provider. We analyze data that has already been audited by the business's internal auditors in order to test whether process mining can identify audit relevant information which warrants further investigation by the internal auditors and that missed detection by them in their review of the same data using their conventional audit techniques. We find that by using process mining we indeed are able to identify numerous instances of such audit relevant information, including payments made without approval, violations of segregation of duty controls, and violations of company specific internal procedures.
Keywords Process mining, internal auditing, ICFR.

* We thank seminar participants at the 2011 International Symposium on Accounting Information Systems in Rome for helpful comments. Further comments are welcome and may be addressed to Mieke.jans@uhasselt.be.

1. Introduction

As the demands by regulators and investors for auditors, both internal and external, to be more effective increase—as evidenced, for instance, by the passage of SAS 99 requiring explicit attention by auditors to detect fraud, as well as the criticisms made of auditors for failing to identify problems with banks prior to the financial crisis—having more and better auditing techniques becomes ever more essential.

Jans, Alles and Vasarhelyi (2010; henceforth Jans et al. 2010) argued that the data analysis technique of process mining has the potential to be a valuable new addition to the toolkit of internal auditors.[1] The concept of process mining was first introduced by Agrawal et al. (1998), little over a decade ago. Since that time process mining has been the subject of extensive research in a variety of disciplines, ranging from engineering and statistics to computer science, and with applications ranging from healthcare and retailing to software development (Bozkaya, Gabriels, and van der Werf 2009; de Medeiros, Weijters, and Aalst 2006; Folino et al. 2009; Greco et al. 2006; Gunther and van der Aalst 2007; Rozinat and van der Aalst 2008; van der Aalst, Schonenberg, and Song 2011; van der Aalst et al. 2003; van Dongen et al. 2005).[2] Process mining is also the subject of intensive interest in business, with the involvement of such leading firms as IBM, Philips and SAP®.[3]

Despite these developments, though, process mining has yet to be used in auditing and this paper is the first to apply the technique to an actual business and to systematically assess whether process mining will prove as valuable in this domain as it has in many others.

The data analyzed by process mining is the *event log* that the auditor constructs from the records maintained by a business's information systems and Jans et al. (2010) drew particular attention to the fact that the event log consists not only of data entered by the auditee, but also what they call *meta-data*, meaning data that is recorded automatically and independently of the person whose behavior is the subject of the audit. Both the data being analyzed and the techniques utilized go beyond current audit practice, which is what made Jans et al. (2010) propose that process mining might identify anomalous transactions that differ from what can be found with existing audit techniques. As they stated in their conclusion:

---

[1] As with Jans et al. (2010), we focus in this paper on internal auditors on the argument that it is more feasible for internal auditors to obtain access on an ongoing basis to the firm's event logs than it would be for external auditors. However, note that in the case of Sarbanes Oxley Section 404 attestations are concerned, the external auditor explicitly relies on the work performed by the client's internal auditor.

[2] A comprehensive list of process mining papers is maintained at http://bpmcenter.org/reports.

[3] http://bpmcenter.org/industry.

> *"The added value of keeping and analyzing an event log lies in the insights into the process that are made possible and in the possibility to detect anomalies otherwise not detectable… by mining the event log one can learn and monitor the process at hand. New insights about the process preceding the final input (like an invoice being paid) are possible. Most important of all for auditing, there are anomalies or frauds that cannot be captured by analyzing input data alone. For instance when procedures leave room for flexibility in order to be operationally efficient, this flexibility can be misused, leading to windows of opportunity to commit fraud or leading to additional indirect costs for the company. On the other hand, too narrowly constrained systems are a major drain on corporate client responsiveness. As we have argued in this paper, frauds brought about by such circumvented procedures may be detectable by process mining the event log data."*

In this paper we explore the value added that the techniques of process mining can provide to internal auditors by conducting a case study of the procurement process at a major European financial services provider. Critically, we analyze not just a set of transactions drawn from a real business but a set of transactions that have already been audited by the business's own internal auditors. Moreover, we worked with the full cooperation of the business's internal auditors, thus giving us the ability to have follow-up investigations carried out on transactions detected by process mining.

As a consequence we are in a unique position to answer the question Jans et al. (2010) pose about the potential usefulness of process mining to internal auditors. Our benchmark for the value added by process mining is identifying what we call *audit relevant information,* which is an anomalous transaction or relationship which warrants further investigation by the internal auditors and that missed detection by them in their review of the same data using their conventional audit techniques.

We find that using process mining we indeed are able to identify numerous instances of such audit relevant information, including payments made without approval, violations of segregation of duty controls, and violations of company specific internal procedures. Whether what we find are indications of fraud, or failed controls, or quite legitimate transactions is a question that only the business's own internal auditors can answer based upon their follow up inquiries. But the outcomes of those investigations—which, for reasons of confidentiality, are not always revealed to us—are immaterial as far as the objectives of the paper are concerned, which is to show that process mining does indeed identify audit relevant information over and above those obtained using only conventional audit techniques.

Looking in detail at the process analysis conducted in this paper, we used four of what Jans et al. (2010) called the different *tasks* in process mining: 1. process discovery, 2. role analysis, 3.

verification by attribute analysis, and 4. social networks analysis. In the first task we focus on discovering how the procurement process actually takes place in practice, in contrast to how it should in the ideal. In the second task we analyze the roles of persons in the procurement process by testing segregation of duty controls. In the third task we use the attributes of cases in order to verify the efficacy of the business's internal controls over the procurement process. In the fourth task we search for social networks among employees involved with the procurement process in order to better understand the interaction between roles and activities for anomalous transactions.

While the generality of our conclusions is limited by the fact that we are necessarily restricted to a limited process view of a single, albeit large and important, business, our paper is an essential follow up to Jans et al. (2010) argument that process mining be explored as a promising new tool for internal auditing. Given the success of this case study, further work by both academic researchers and practitioners is clearly called for in order to better delineate the specific circumstances under which process mining can provide the greatest value added.

The remainder of the paper is organized as follows. In the next section, a brief background is provided on the role of business processes in internal auditing and on process mining. Section 3 presents an overview of the case study. Section 4 then discusses the procurement process, which facilitates the creation of the event log, which is discussed in section 5. Section 6 reports the results of the process mining analysis using four different tasks and compares the findings against those of the standard internal audit of the same data. Section 7 offers concluding comments.

2. Processes in Auditing

2.1 Business Processes and Internal Controls

A process focused approach in internal auditing has been discussed in numerous papers over the years (Kopp and O'Donnell 2005; O'Donnell and Schultz Jr 2003; Bierstaker, Hunton, and Thibodeau 2009; Carnaghan 2006). These papers deal with such issues as the use of business-process-focused audit support software and internal control evaluation, client prepared flowcharts and process modeling language. As discussed above, Jans et al. (2010) brought process mining into the auditing research realm.

The role of business processes has also come increasingly to the forefront of auditing practice since the enactment of the Sarbanes-Oxley Act of 2002 in the United States placed an emphasis on internal control monitoring and reporting. Section 404 of the Act required

management to report on the effectiveness of the business's internal controls over its financial reporting process (ICFR) and for the external auditor to provide an attestation of that management report. In practice, that has meant that internal auditors are heavily involved with the development and testing of ICFR.

Such internal controls do not, however, guard a business against all risks associated with its financial reporting process. Although processes are mostly prescribed to take place according to a designed process model, some room for flexibility is necessarily built in to the control system. This to make allowance for the numerous deviations from the designed process model that are inevitably required in practice for the smooth operation of the business, constraints which would otherwise interrupt the process flow on too frequent a basis. For example, while ideally there would be a three way match between a purchase order, a goods receipt and an invoice, allowance may be made to accept deliveries that include unanticipated transportation costs that were not included in the original purchase order. Otherwise, such deliveries may be rejected, resulting in unacceptable delays to the downstream production process.

In today's heavily IT-enabled businesses, this flexibility is built into their Enterprise Resource Planning (ERP) systems (the most common of which is SAP®, as it is in our case study), and while IT internal auditors periodically check that the settings of the ERP systems are as they should be, they also anticipate that over time setting will change to allow for exceptional transactions, cope with changes in personnel and so forth (Alles et al. 2006). Hence, the reality is that it is not always feasible to rigidly "lock down" internal controls, implying that only monitoring ICFR is not sufficient to cover all the risks associated with the business process. The auditor also has to anticipate that there will be deviations from the ideal process and will have to test in detail to see whether such deviations are acceptable, or evidence of control failure.

It is in this context what we argue that the use of process mining techniques give internal auditors a new tool to assess the efficacy of ICFR.

2.2 Process Mining

Jans et al. (2010) provide an extended discussion of process mining and we only summarize that material here. The Business Process Management Center describes process mining in the following terms: [4]

> *"The basic idea of process mining is to extract knowledge from event logs recorded by an information system. Until recently, the information in these event logs was rarely used to analyze the underlying processes. Process mining aims at improving this by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs. Fuelled by the omnipresence of event logs in transactional information systems […] process mining has become a vivid research area."*

As the quote indicates, the source of data for process mining is an *event log*, often referred to as *history, audit trail, transaction log*, etc. (van der Aalst et al. 2007).[5] Most businesses of any significant size today store their data, including log data, electronically thanks to the maturing of technologies for databases and computer networks, particularly the ubiquity of ERP systems in businesses of any significant size. Process mining is a term subsuming all methods of distilling structured process descriptions from a set of real executions, using event log data. (van der Aalst, Weijters, and Maruster 2004)

The event log that is used in process mining is constructed from raw data extracted from the business's information system, almost always its ERP system. There are four characteristics that need to be extracted from the information system about each event in order to facilitate any kind of process mining analysis: 1) the *activity* taking place during the event, 2) the *case* or *process instance* of the event (for example, an invoice), 3) the *originator*, the party responsible for the event, and 4) the *timestamp* of the event. For example, the event with unique identifier 000001 refers to a *Sign* (activity) of *purchase order 4603* (case) by *Ann Smith* (originator) on *February 5th 2011* (timestamp). Critically, at least some of these four characteristics are logged by the IS independent of the originator, meaning that in this example, Ann Smith is required to log into the ERP system with her unique password before being able to enter the transaction, and the system routinely timestamps the data entry, with Smith not having the user privileges that would enable her to override or modify that automated logging. Collecting and systematically organizing information on at least these

[4] The BPM Center is a collaboration between the Information Systems groups (IS@CS and IS@IEIS) at Eindhoven University of Technology (Eindhoven, Netherlands) and the BPM group at the Faculty of Information Technology of Queensland University of Technology (Brisbane, Australia).

[5] Note that the term "audit trail" comes from computer science and not from accounting. Despite the name, there is little use made at present by actual auditors of audit trails.

four characteristics, as well as incorporating any other available attributes of the transactions, results in an event log which can be analyzed using process mining techniques.

3. Case Study Overview

The case study is undertaken at a leading European financial service provider which ranks among the top 25 banks in the world (and which is also subject to the provisions of the Sarbanes Oxley Act because of its operations in the United States). The procurement process is the focus of the study, chosen because it is a typical, standardized business process in most businesses around the world, one which represents a large expense item for even a non-manufacturing business (1.4 billion Euros in this case) and with implications for the financial reporting process. Indeed, many ICFR controls focus on the procurement process, such as segregation of duty controls.

The method of our analysis is to extract from the business's SAP® system a set of transactions from which we could construct a complete and coherent event log which could then be analyzed using process mining techniques. Hence, the first step is to select an appropriate data set.

An internal auditor undertaking process mining in practice would focus on the most recent transactions of the business, but for the purposes of our research, timeliness is less important than having data that is already audited, so that we have a benchmark with which to assess the value added of process mining. Hence, the primary criteria used in the data selection was that the data had to have been specifically audited by the business's internal auditors, and that the full year concerned had to have had an unqualified opinion from the external auditor, as well as a clean Sarbanes Oxley Section 404 statement from management about the efficacy of the business's ICFR over that process, accompanied by a clean attestation of that statement from the external auditor. In addition, given that the case study concerns a bank, we felt it best that the data set for the process analysis precede the onset of the financial crisis in 2008.

Given this selection criteria, the choice we made for the transactions that are the subject of our case study were the set of invoices paid during the month of January 2007. These invoices are then traced back to their corresponding purchase orders (POs), which were issued between 2005 and 2007. It is these POs which represent the cases we then follow throughout the procurement process when creating the event log and process mining it. Note that we did not take a sample of this data as conventional audit practices would. Rather, our process mining techniques allow us to analyze the entire population for the relevant transactions (i.e. all

invoices paid in January 2007), thereby providing evidence for the applicability of process mining techniques to run on a population set on a monthly basis.

While our data set is not current, the advantages of using already audited data outweighs the disadvantages, if any, of its timeliness, since the age of the data is largely unrelated to the analysis it is subject to—purchasing is a standard and ongoing business activity and it has not changed significantly since the time the data is drawn from, and nor has the way in which the internal auditors conduct their standard audit of it. Moreover, by using non-contemporaneous data, the business's internal auditors were in a position to then examine their records and report back to us on the nature and meaning of the audit relevant information that we detected using process mining.

Data is available for the construction of event logs because the business uses SAP® as its ERP system, including for their procurement cycle. The configuration settings of their ERP system are the focus of the ICFR as the business aims to ensure that its procurement process follows established guidelines. As mentioned above, the procurement department was the subject of a standard internal audit before we applied process mining to data drawn from the same time period covered by the audit. This provides our benchmark for the value added by process mining over and above conventional internal audit practices.

Before creating an event log and process mining it, the auditor has to identify which activities constitute the business process. Thus the first step in process mining is to undertake a process analysis, in which the precise details of the business process are systematically gathered and a process map created. A process map consists of four components which describe the business process: 1) process objectives, 2) activities, 3) information flows, and 4) the accounting impact. (Knechel 2001) For the purpose of process mining, the second component is the main focus, as it is the principal input for the creation of the event log.

A variety of research tools were used to obtain the necessary details about the procurement process at the business. Executive officers (both business and information systems specialists) were interviewed, employees at various departments were questioned and observed during their job, and the internal user guidelines of the ERP system were consulted.[6]

[6] The information gathered during the process analysis is not reported as a full blown process map since constructing that is only a means towards an end in this case study. For reasons of simplicity and comprehensibility, only a process flow chart is presented rather than a full description of all four components of the business process.

After the process analysis is executed, an appropriate event log is built from the stored log data. The relevant log data that is captured by the ERP system is vast in magnitude and dispersed over numerous tables of the SAP® system, a system whose logic schema is partly specific to the ERP system, and partly business specific. In order to mine this data it needs to be systematically configured into an event log with the format requirements that facilitate process mining. This format structures the relevant data around the activities that constitute the process. As Jans et al. (2010) emphasize, creating an event log is not something that is automated at present, and represents a significant commitment on the part of the researcher even before process mining can be undertaken.

Once the event log is created, the process is analyzed by applying the four different tasks outlined above: analyzing the order of activities to discover how the procurement process is carried out in practice; analyzing how activities are linked to persons through a role analysis; a verification phase consisting of analyses to further examine the output of the previous steps; and social network analysis. We can then determine whether the anomalous events that are uncovered during the process analysis were also detected by the internal auditors in their standard analysis, or represent newly discovered audit relevant information.

4. Activity Flows in the Procurement Process

The flow chart in Figure 1 provides an outline of the main activities in the procurement process. The process is triggered by the creation of a purchase order (PO) by an employee. This PO has to be signed and released by two distinct persons, thereby approving the order for release. Once the release is approved, the employee can order the goods from the supplier.[7] The supplier will then dispatch the goods and the accompanying invoice. If both the documents 'Goods Receipt' (GR) and 'Invoice Receipt' (IR) are entered into the system, the accounting department will book the invoice. This last activity will trigger a payment.[8]

Figure 1 illustrates only the ideal or *designed* process model. In practice, the process can deviate from this designed model. For example, changes can be made to the PO after its creation, perhaps triggering a new 'sign' and 'release' activity. This would result in an extra activity that is not in the process model ('change PO') plus an extra arrow that redirects to the activities 'sign' and 'release'. Another example is the receipt of goods in multiple deliveries.

[7] Since this takes place outside the ERP system the supplier activity is not depicted in Figure 1.
[8] The payment occurs automatically and a payment record is stored in another information system and is for this reason not seen as a different activity from the 'book invoice' activity. We therefore depicted activity 6 as "Pay" instead of "book invoice".

This would cause extra activities of ‘receive goods’, perhaps followed by multiple invoices, and hence extra activities of ‘receive invoice’.

In theory, the ERP system can be configured in such a way that such deviations are not allowed and, hence, become impossible to execute. But locking down the process in this way would result in a constant stream of exceptions and delays since the actual procurement cycle would often deviate from the designed process for a variety of reasons, some anticipated and acceptable, and some not. For example, there could be problems in manufacturing the items ordered or shortages in their availability, thus resulting in partial shipments. All procurement processes have to build in flexibility to cope with such issues for otherwise the system would either fail, or be sidelined by employees, thus creating a bigger control problem.

Hence, in practice process executions can deviate frequently from the designed process. This is the real challenge facing the auditor: once primary controls are relaxed allowing deviations from the designed process, how can control be retained over the process? Aside from monitoring the installed controls, the auditor has to analyze the deviations from the designed model, bearing in mind that not all exceptions are necessarily indications of internal control failures. Some process deviations are normal, others are suboptimal and others are indeed anomalous outliers that at a minimum require further investigation. Hence tests of details are required to supplement the tests of controls.

Having established the broad outline of the procurement process, we can turn to the creation of the event log.

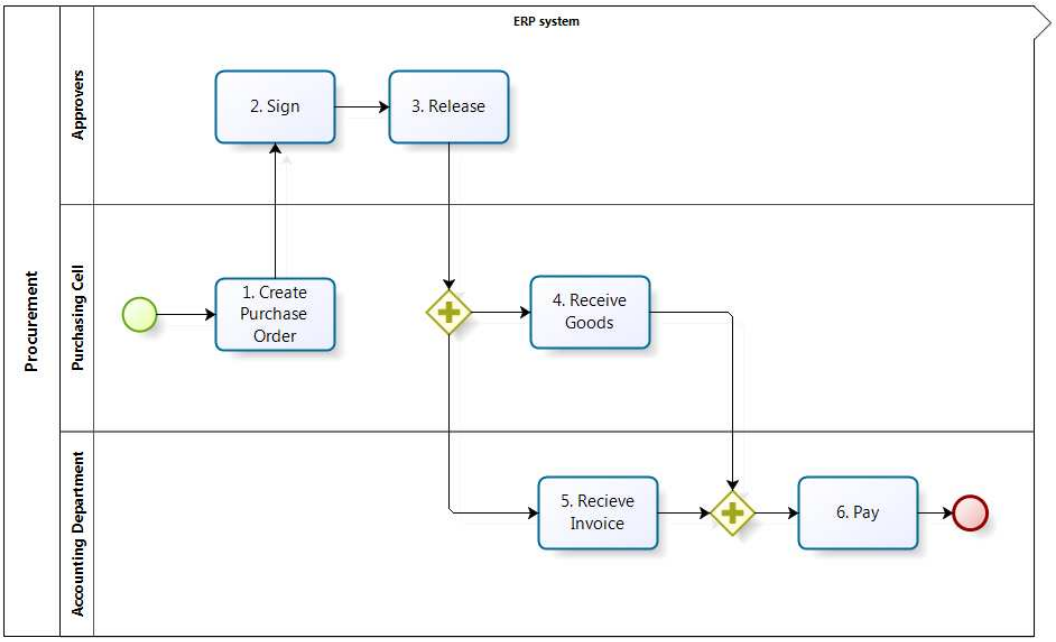

Figure 1: Flow chart overview of the procurement process at the case company

5. Creating the Event Log

In order to begin creating the event log two preparatory questions need to be resolved: 1) what are the *activities* that constitute the process? And 2) what is the *case* or *process instance* that is the focus of these key activities? These questions are important because it defines the content and structure of the event log and their choice is a key judgment call by the auditor. The auditor has to use the output of the process analysis to identify the key activities in the process. Then it has to be determined what cases move through these key activities and whether it is possible to collect data on these cases. If the auditor does not have a thorough understanding of the business process, this will reduce the power of the mining results since the discovered model from the event log will fail to include all relevant activities.

The activities we select to include in the event log in this case study are based on the information gathered during the process analysis step. The six activities, represented by the six rectangles in Figure 1 are selected to incorporate into the event log along with the additional activity of “change” to the original process. This activity is not represented in the designed process model (and hence, not shown in Figure 1), but is available for execution in the ERP system as a means of building in flexibility into the business process. Relevant data on these seven activities (timestamp and originator) was located in the SAP system, assuring the necessary data availability to proceed with the event log creation.

Aside from the selection of activities, another key decision to be made by the auditor is the choice of the process instance. A process instance is a case or subject that can be systematically followed throughout the process by linking its unique identifier to the activities that it undergoes.

For our case study we chose as our process instance the purchase order line item. We preferred the detailed level of a PO item line over a PO as a whole, because the booking of an invoice in the financial ledger is based on each item line. So although a PO is approved as a whole and not per line, we followed the PO lines individually throughout the process from being created to being paid. Having made this selection, the activities we selected need to be related to this process instance choice. Note, however, that the activities 'create PO', 'Sign', and 'Release' refer to the parent PO that the item line belongs to, since there is no timestamp available on the creation of an item line, and signs and releases are executed on a PO level as a whole.

Based on choice of the activities and the process instance, all relevant logged data concerning these activities is retrieved from the various tables of the SAP® system and configured in an event log format. The software we used for the process mining itself is ProM, which is an established open source framework for executing process mining tasks.[9] In order to apply process mining techniques using the ProM framework the event log needs to be in the MXML format (Mining XML). As a consequence, we extracted the data from the ERP system, manipulated it in SAS®, exported it into MS Access® and then used the ProM Import tool to convert the Access database into an MXML file.[10]

As explained above, we need four characteristics in order to mine a process: an activity, a case, an originator, and a timestamp. Once the first two have been chosen, we are able to state for each line item of a PO whether, when and by whom it was created, signed, released, the goods and invoices received, the payment completed and a change executed. The timestamp of when these activities were executed and the originator record of by whom, are meta-data, being logged by the SAP® system automatically and are not based on data input by the originator.

By contrast, there is in addition to these four fundamental characteristics, other information entered by the employee relating to the process instance. These are thus *input-data*, as

[9] http://www.win.tue.nl/processmining/prom/start.

[10] More details on how the MXML format is structured and how our procurement data was converted into MXML can be found in Jans et al. (2010b).

opposed to the meta-data, and are called *attributes.*[11] These attributes are an essential component of the event log, including, for example, the value of a PO, the delivery address, the document type of a PO, the reference number of an invoice, and the reference number of the Goods Receipt document the invoice refers to. These are only some of the details of transactions that are recorded within the SAP® system and which are available for incorporation into the event log, though the auditor faces a tradeoff between the number of attributes included in an event log and the difficulty in creating and analyzing it.

For each activity the timestamp and originator are extracted from the ERP system and a link is made to the process instance (PO item line) so that an activity flow, called an *audit trail*, is stored in the event log for each process instance under investigation. An audit trail of a case can, for example, be as follows: *Create PO → Sign → Release → Goods Received → Change Line → Invoice Received → Pay.* Aside from information on the timestamp and originator, extra attributes are also stored. The attributes are divided into two groups. There are attributes of the process instance (such as the value of the PO item line, the purchasing group it belongs to, etc.) and attributes of the activities, the latter attributes depending on the activity it relates to. For instance an attribute of the activity 'Change Line' is the modification that took place (in the case of a changed PO, its new value; otherwise zero), while for the activity 'Goods Receipt', the attributes include the value and the quantity of the goods received and a reference number of the accompanying invoice.

As mentioned above, our event log contains all invoices paid in January 2007. The invoices were then traced back to their accompanying PO's—we found no invoices without POs in this population. These PO's, with creation dates between 2005 and 2007, were then followed from their start activity 'Create PO' to their end activity of payment. If the end activity was not a single payment but rather multiple payments, then the ending activities were cut off at the last payment activity, with this last payment taking place in January 2007. Hence, our event log only contains completed procurement cycles. Obviously this choice restricts the kinds of anomalies that can be detected, but then again, it prevents false alarms from being generated by POs without a payment yet in the system because the procurement cycle is not yet completed.

Our event log consists of 26,185 process instances (i.e. PO lines), involving 181,845 activities and 272 originators. The average audit trail consisted of six events, with a minimum of four

[11] In this case study all attributes were input-data, though that need not be the case in general.

and a maximum of 390 events.[12] The frequency of activities in the event log is summarized in Table 1. The number of cases in our event log (26,185) has to equal the activity of 'Create PO', since this activity refers to the creation of the parent PO that this PO item line belongs to. This activity occurs exactly once per case. If the ideal procurement process is followed, then all activities should take place the same number of times. The fact that they differ is immediate evidence that the actual process differs considerably from the designed process, indicating either a necessary flexibility to accommodate business needs, or a failure in ICFR.

From Table 1 it can be seen that there are less Sign activities than PO (lines) created, meaning that not every PO is signed. On the other hand are there more releases than PO lines. We also identify more payments than cases, implying multiple payments on one PO line.

Activity	Number	Activity/Create PO
1.Create PO	26 185	1.0000
2. Sign	25 648	0.9795
3. Release	28 748	1.0979
4. GR	24 724	0.9442
5. IR	29 255	1.1172
6. Pay	31 817	1.2151
Change Line	15 468	0.5907

Table 1: Frequency of activities in the event log

In summary, our event log consists of POs issued between 2005 and 2007 and which were paid in January 2007. But while we know the starting point and end point of these transactions, what is of interest to the auditor is how the process went from one to the other. Discovering that process and analyzing its details is what process mining of the event log enables us to do.

6. Results of the Process Mining Analysis

6.1 Discovery Analysis of the Procurement Process

The most fundamental use of process mining is to analyse the event log in order to discover how the business process is actually carried out, as contrasted with the ideal designed process model, from which deviations have taken place in practice. Process discovery is carried out by examining timestamps to systematically establish the flow of activities of each PO line, from creation to payment. This type of analysis is unique to process mining, since it utilizes the

[12] This maximum audit trail is likely an open order—one where a single PO line is used over and over again—but nonetheless, it is clearly a transaction that warrants further investigation by the internal auditors.

meta-data on activities and timestamps. Using traditional analysis techniques would not yield these insights.

In our case the Performance Sequence Analysis reveals 304 distinct patterns, with the six most frequent patterns shown in Table 2. This is certainly more than the one pattern that comprises the ideal designed procurement process, but whether 304 patterns is more or less than what one might expect of a business of this size and complexity is an open question. Unfortunately the fact that process mining is such a new audit tool means that we lack a knowledge base drawn from procurement and other business processes from a variety of businesses that can serve as a benchmark. Even in the absence of such a basis of comparison we can confidently state that most internal auditors would be surprised that there are such a variety of ways in which the procurement process is being carried out, and moreover, that the patterns vary to the degree that they do.

On the other hand—given the ubiquity of 80/20 phenomena in business and elsewhere—it is probably less surprising that as Table 2 indicates, just three out of 304 patterns cover over 80% of the data set. By contrast, there are 104 patterns that only occur once.

Relating the six patterns to the designed model in Figure 1, we recognize in pattern 1 the designed procurement process, with no changes or deviations. However, well under half of all transactions correspond to this ideal. Pattern 2 differs from the ideal in that there is a change that takes place in the PO after it is created, but that is not a major audit concern by itself since the change takes place before the PO is approved.

By contrast, in patterns 3 and 4 the Sign activity is absent altogether, contrary to the specifications of the designed model. When these patterns were brought to the attention of the internal auditors, they determined, after subsequent investigation, that it was in fact legitimate to release these particular POs without a signature due to the specific nature and size of these transactions. Those specific circumstances related to particular document types and associated thresholds, similar to the way in which some senior managers can release POs concerning payments below a threshold level on their own authority.

In patterns 3, 4 and 5 there is no Goods Receipt document entered into the ERP system. This can happen quite legitimately when the “good” in question is in reality a service for which there is no act of delivery to a shipping dock. In fact, in the case of such services the goods receipt indicator is supposed to be flagged off in the system to indicate that no GR entry is to be expected, but possibly this action is overlooked by originators. Clearly an auditor would

likely want to investigate these patterns to ensure that they do indeed represent services rather than goods, and that the services purchased are appropriate to the business. In the last pattern, the IR and GR have switched their expected order. As depicted in Figure 1, though, these activities are allowed to appear in parallel order which would explain this phenomenon.

If several of just these six non-ideal patterns in Table 2 warrant some sort of further examination by internal auditors, then what of the remaining 298 patterns? One the one hand, one can dismiss them as non-material, encompassing as they do, less than 14% of all POs. On the other hand, that may be precisely why it is they that should be of the greatest concern to the internal auditors. As the audit standard AU 329 states, auditors are required to undertake analytic procedures in order to find the outliers that may indicate anomalies or fraud.[13]

Unfortunately we are unable to pursue this line of inquiry any further in this limited case study. As before, we are handicapped by the absence of benchmarks that would enable us to match the patterns in our procurement process against a known set of anomalous patterns. But the findings of this case study indicate why creating such a knowledge base through the more widespread application of process mining and dissemination of its results would be of value to auditors.

		Pattern Frequency		Cumulative total	throughput time (days)			
Pattern	**Sequence**	**#**	**%**	**%**	**avg**	**min**	**max**	**st.dev**
1	Create PO → Sign → Release → GR → IR → Pay	11,608	44.3%	44.3%	27.78	1	334	20.05
2	Create PO → Change Line → Sign → Release → GR → IR → Pay	6,955	26.6%	70.9%	32.33	2	343	57.72
3	Create PO → Change Line → Release → IR → Pay	2,488	9.5%	80.4%	75.63	3	344	38.99
4	Create PO → Release → IR → Pay	640	2.4%	82.8%	16.8	3	338	26.38
5	Create PO → Change Line → Sign → Release → IR → Pay	491	1.9%	84.7%	50.85	6	237	24.07
6	Create PO → Change Line → Sign → Release → IR → GR → Pay	393	1.5%	86.2%	56.36	9	295	40.16

Table 2: Most frequent patterns in the event log

In the absence of a suitable benchmark, we proceed in our analysis of the discovered process patterns by comparing them against each other by using a process discovery algorithm to analyze the sequence of activities within these patterns. Given the large number of patterns,

[13] *"A basic premise underlying the application of analytical procedures is that plausible relationships among data may reasonably be expected to exist and continue in the absence of known conditions to the contrary. Particular conditions that can cause variations in these relationships include, for example, specific unusual transactions or events, accounting changes, business changes, random fluctuations, or misstatements."* AU 329 (http://pcaobus.org/Standards/Auditing/Pages/AU329.aspx)

some with a large number of activities (up to a maximum, recall, of 309 activities), simple observation no longer suffices to analyze the patterns, hence the use of a systematic algorithm to better visualize them.

To begin with, we apply the Fuzzy Miner algorithm of Günther and van der Aalst (2007), using its default settings. This algorithm filters for the typical issues encountered with large data sets such as completeness and noise and then simplifies and visualizes complex processes. The output is depicted in Figure 2. The thicker the line, the more frequently a sequence of activities occurs. The core process shown corresponds to the designed process, as one might expect, given the frequencies in Table 2. The deviations are a Change Line that often occurs between the creation of the PO and the Sign as in pattern 2 in Table 2, and we note the existence of loops on every activity but the creation. There is also some interaction between the payment and the invoice receipt, which was established to be legitimate after discussion with the process owners and the internal auditors.

Using the default settings in this analysis only reveals the core sequences in the event log. To uncover the less frequently followed sequences, we set lower thresholds of the metrics in the algorithm, resulting in the model in Figure 3.

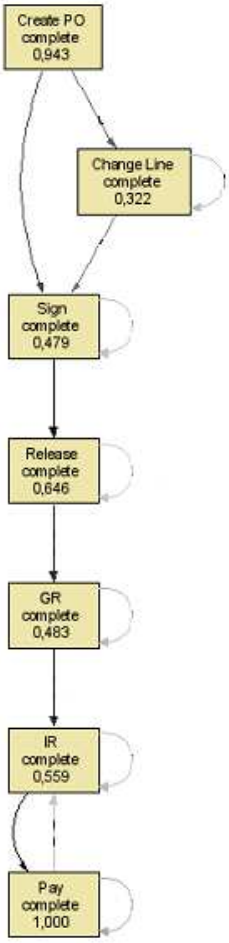

Figure 2. Output of Fuzzy Miner with default settings to uncover the core process in the event log.

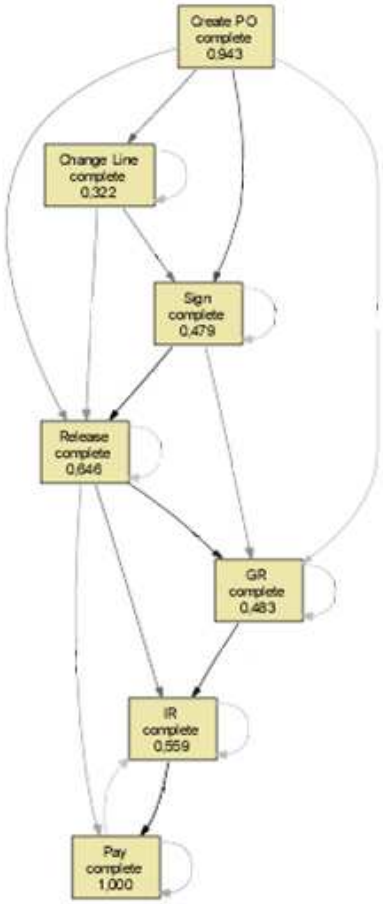

Figure 3. Output of Fuzzy Miner with lower threshold settings to reveal less-frequent flows

In Figure 3 we see a more complex process with extra flows (*edges*). However, we have to be careful with interpreting these extra flows, because there could be an AND or an OR relationship behind an edge. For instance, there could be a particular flow like Sign → IR depicted while in fact this is part of an AND relationship like 'after Sign: Release AND IR occur'. Further, the fact of having extra flows does not, by itself, indicate that there is a control failure or fraud taking place, but such transactions need to be examined in a verification phase.

In order to better identify the flows that require further investigation, we specifically check using a Linear Temporal Logic algorithm whether the extra flows depicted above also really prevail in this order. This check is performed on the whole population set, exploiting the meta-data on activities and timestamps. The results of the checks are summarized in Table 3.

Notice that the flows shown in Table 3 are only subsets of complete audit trails, and unlike the entries in Table 2, they do not show a complete pattern.

We find that only the sequence Create PO → GR does not occur in the event log but that all other conceivable flows do in some part of the various patterns. For example, flows 2 and 3 are absent a Sign activity. Upon questioning of our process experts, it was established that there are situations where a release alone is sufficient for approval, but only when additional conditions are met (maximum amount and specific document type). Whether these conditions were met in these cases cannot be established using the process discovery toolset, but will be examined below in the verification phase of process mining.

There are also eleven cases where a Sign was immediately followed by a Goods Receipt. This is in violation of ICFR where a GR can only take place after a release. While these cases were subsequently cleared by further investigation, the process mining was clearly successful in discovering flows that did need to be investigated by the internal auditors. The Sign activities which occur at PO header level and not at the detailed process instance level were all triggered by a change in another item line than the process instance itself. The GR activity on the other hand is related to the process instance itself and was not associated with the Sign that took place just before the GR. The flows Release → IR and Release → Pay both stress the importance of the Goods Receipt indicator. The business makes it permissible to discard the GR activity, but the Goods Receipt indicator needs to be flagged off in that case. Whether this procedure was complied with will also be examined during the verification phase. Finally, flows 6 and 7 stress the importance of examining whether there exists for each payment a corresponding invoice. This too will be checked in the verification phase.

	Extra flows	**Occurrences**	**Result**
1	Create PO → GR	0	OK
2	Create PO → Release	739	Verification required for omitting Sign
3	Change Line → Release	2.790	Verification required for omitting Sign
4	Sign →GR	11	Further investigation → OK
5	Release → IR	4.973	Verification required on GR indicator
6	Release → Pay	244	Verification required on GR indicator and IR
7	Pay → IR	227	Verification required on IR

Table 3. Results of explicit checks on the extra edges in the Fuzzy Miner output

As a last, very rudimentary control, we check whether each PO process instance has at least one release activity in its pattern. We find three cases out of the population of 26,185 cases where there is no release. On two occasions that is due to a process instance that was created by a batch file, was paid and subsequently reversed. Somehow these two transactions got

through the system without an approval. In the third case further investigation revealed that the approval has been taken place outside the SAP workflow which is why it was not in recorded event log.

There is more analysis that could be conducted of these patterns, incorporating, for instance, the data shown on Table 2 of their throughput time. But the aim of this paper is not the comprehensive examination of the event log, but rather, to establish that there is value added to process mining. Clearly the process discovery task revealed numerous examples of audit relevant information which warranted further investigation by the internal auditors, which is support for our assertion. We now turn to the other process mining tasks to better define the potential contribution of process mining to internal auditing.

6.2 Role Analysis

Role analysis exploits the presence of meta-data on activities and originators in the event log to examine the part played by employees in the procurement process. Particularly following the passage of the Sarbanes Oxley Act in 2002, businesses have invested heavily on preventive controls on segregation of duties (SOD), in order to ensure that the same individual is not responsible for all critical steps in a process, such as both creating and signing POs.

On the other hand, since an individual often executes several activities, they can have multiple roles in overlapping processes, and in addition, the inherent flexibility in ERP systems can lead to slippages in control over time as personnel change their employment status or their roles. Hence there is a need not just for tests of controls, but also tests of detail for the roles actually played by employees in carrying out the procurement process.

In our case study, three fundamental SOD controls are meant to be followed in the procurement cycle:

1. The Sign and Release activities for a given PO should be undertaken by two distinct individuals.
2. The Goods Receipt and Invoice Receipt activities for a given PO should be undertaken by two distinct individuals.
3. The Release and Goods Receipt activities for a given PO should be undertaken by two distinct individuals.

As the first step in undertaking the role analysis, we created an Originator-Task matrix from the event log which details the number of times an individual executes a particular activity.

From this matrix we can conduct a preliminary check as to whether some individuals execute an impermissible double role. With 272 originators in this case study, the full table is too large to show, but from the excerpt presented in Table 4 we find, for example, that individual '…1' undertakes both the Sign and the Release activity. Similarly, individual '…4' undertakes both Goods Receipts and releases. No example of an individual combining the GR and IR roles is found in the matrix. It is important to note, however, that the matrix is a very preliminary analysis in that it only shows total activities, and not activities isolated by the process instance. Thus the 11 cases which individual '…1' released may or may not coincide with the 171 that he or she signed, and there is obviously no violation of SOD controls if they do not overlap. The situation could simply reflect a reassignment of responsibilities, perhaps due to a promotion of the individual involved, or the need to temporarily replace an absent colleague in the Sign role. Hence, identifying individuals with combined roles only highlights audit relevant information that warrant further investigation by the internal auditors, but by using the tools of process mining we can also test SOD controls more comprehensively, on a case by case basis.

Originator	1. Create PO	2. Sign	3. Release	4. GR	5. IR	6. Pay	Change Line
….1	0	171	11	0	0	0	0
….2	0	0	0	0	280	310	0
….3	0	0	23	0	0	0	0
….4	0	0	42	42	0	0	0
….5	0	24	0	0	0	0	0
….6	152	0	0	189	0	0	204
….7	0	0	10	0	0	0	0
….8	0	0	66	0	0	0	0
….9	0	0	1	0	0	0	0
….10	207	241	0	199	0	0	155
….11	0	0	0	15	0	0	11
….12	4572	259	0	4517	0	0	244
….13	…	…	…	…	…	…	…

Table 4: Excerpt of Originator-Task matrix

Given the size of the Originator-Task matrix, visual inspection is not an appropriate to detect all suspect SOD instances and hence we utilise a Linear Temporal Logic tool to check whether the three fundamental SOD controls hold for each PO.

The first assertion—that Sign and subsequent Release of a PO are by two distinct individuals—needs to be tested pairwise, since there can be multiple signs and releases for one process instance (though this should not happen in the designed process, in reality, as the process discovery showed, there are numerous variations and loops in the actual process). For

instance if a release takes place and then a line is changed, the next Sign is allowed to be performed by the previous releaser. That is not an issue with the other two SOD controls.

After testing the entire population of 26,185 POs we can conclude that the first two SOD controls hold without violation in the investigated event log. Concerning the third assertion, 175 violations were found. Close examination revealed that these exceptions involved only three individuals. One individual violated the SOD control on GR and Release 129 times, another individual incurred 42 violations, while the third individual did so four times.

These 175 cases revealed by the role analysis task are clearly audit relevant information and demonstrates the value this type of process mining can have in internal auditing. This evidence was handed over to the internal auditors for follow up investigation. For reasons of confidentiality, we cannot discuss the outcome of that investigation.

6.3 Verification by Attribute Analysis

The discussion of the two previous process mining tasks indicate that some outcomes need further investigation to assess whether or not they represent violations of controls. Some of this investigation has to be undertaken manually by the internal auditors, but in other cases it can be done through further process mining tasks, by exploiting the information on attributes of the process instances available in the event log.

As mentioned above, an attribute may contain information on the process instance itself or on an activity the process instance is submitted to. The analyses in this section are a direct response to the output of the activity patterns found in the primary analyses, reported in Tables 2 and 3. A first analysis compares the references of the payment activities with the references of the IR documents, to check whether there is an accompanying invoice for each booked payment. Both reference numbers of the payment and the invoice are stored in the event log as attributes. This test resulted in 46 incorrect process instances, encompassing 265 stand-alone payments. One process instance has 131 pay activities without a corresponding IR, another 75, and yet another, 10. The remaining process instances only have one, two or three stand-alone payments. There were 17 originators responsible for these payments. One of these originators is responsible for 216 out of the 265 payments. Two other individuals have respectively 18 and 12 stand-alone payments on their account.

These payments were all investigated (manually) to check whether the payments could have been based on a ‘Subsequent Debit’, which is an acceptable alternative document for a standard invoice. This indeed appears to be the case with all these payments. The question

remains, for follow up investigation, why all these bookings are based on this type of document instead of on a regular invoice.

A second analysis, also as a follow-up of the revealed patterns in the process discovery task, investigates the functioning of the Goods Receipt indicator. If this indicator is flagged on, the accompanying process instance should have a GR before it can be paid. We tested whether all cases without a GR indeed had a Goods Receipt indicator that was turned off. There were three cases where this assertion did not hold, indicating a breach in the configuration settings of the ERP system. As discussed above, in this context it would be useful to have an attribute on whether this case refers to services or goods. That there is no such field in the ERP system could be considered a shortcoming revealed by process mining.

The last attribute analysis verifies whether the internal conditions of the organization are met when there is no Sign in the activity pattern. There were 742 cases (2.8% of the total) which both lacked a Sign activity and failed to meet the conditions under ICFR where such an omission is permissible. This evidence was handed over to the internal auditors for follow up investigation. Once again, for reasons of confidentiality, we cannot discuss the outcome of that investigation.

6.4 Social Network Analysis

The originator entry in the event log allows us to construct a social network of all employees involved in the procurement process. In Figure 4 the social network of all the employees is depicted, with each circle representing one out of the 272 individuals in our population. The range of interactions between such a large numbers of employees results in an output that is difficult to gain much insights from. Where social network analysis is particularly valuable is when it can be focused on a specific subgroup of interest.

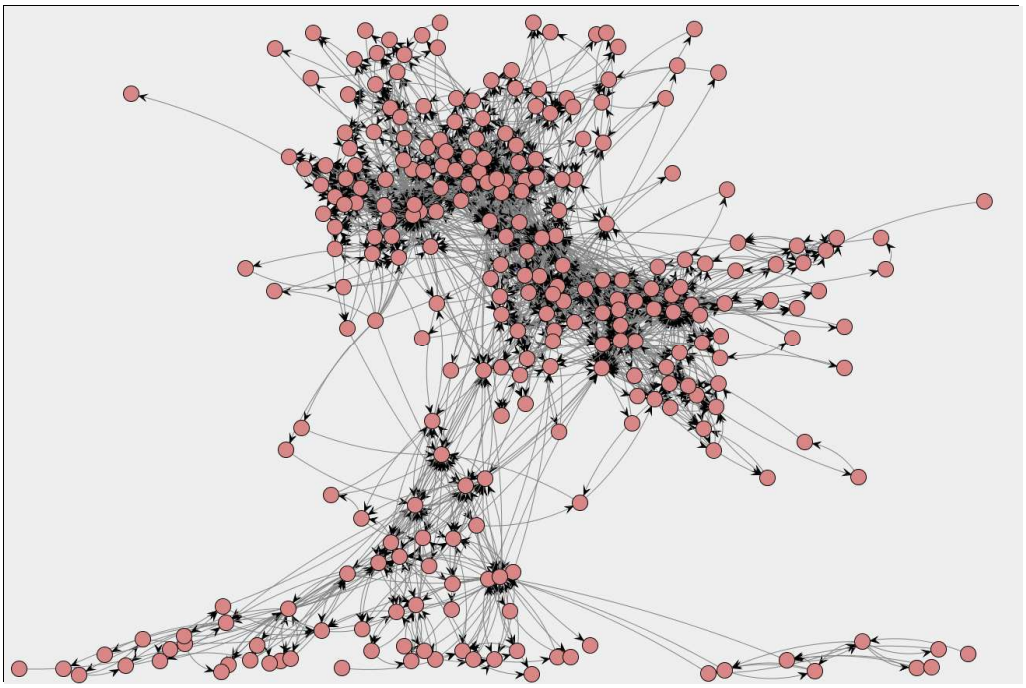

Figure 4: Social Network of all originators in the procurement process

Hence we undertook a social analysis of the subgroup of 175 cases by three individuals that the role analysis revealed were in violation of the SOC control concerning the activities 'release' and 'Goods Receipt' being undertaken by the same individual. The social network analysis centers on the three originators directly responsible for the violation and maps which other employees interacted with them in the event log.

We determined that in total, 21 other individuals were involved with the three primary originators across the 175 cases. The social network of these 24 individuals for these cases is depicted in Figure 5. As can be seen, there are three distinct clusters, with the three individuals violating the segregation of duty controls shown in the central position of each cluster (these three employees are identified by the standalone [red] arrow →). This map of their social networks provides the opportunity to compare the designed organizational structure with the actual network.

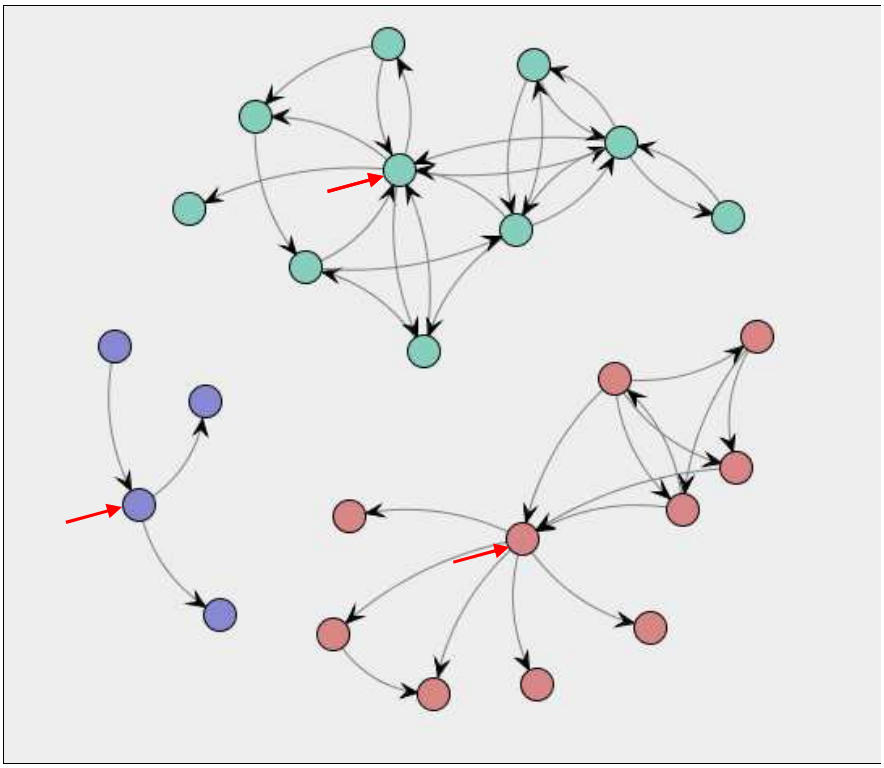

Figure 5: Social Network of 175 cases by three individuals violating SOD controls on Receive and GR

Another interesting subgroup to visualize is the social network of the individuals involved with the 742 cases identified by the attribute analysis where no Sign was present and the conditions for this exception were not met. As before we construct a social network diagram by using the originator entries in the event log cross-referenced against the 742 process instances. We see from the diagram in Figure 6 of this social network that there are three clusters of employees, with two of the clusters connected to each other by two individuals who are involved with both groups. By contrast, the third group is both completely isolated and involves very few individuals.

A social network analysis is not an end in itself, but a means towards obtaining insights into the meaning and motivation of transactions through understanding how the individuals involved relate to each other in an organization. As Jans et al. (2010) speculated, social network analysis may be a way of tackling one of the most intractable problems facing auditors: collusive fraud. The clusters shown in Figures 5 and 6 are not evidence per se of such fraud, but at least it limits the scope of the follow up internal audit investigation to a manageable subset of all employees. The real value added comes when, as we have done here, social network analysis is combined with other process mining tasks, such as process

discovery and role analysis. This exploits the full dimensionality of the data in the event log and focuses attention on the most serious violations of controls.

The evidence obtained from the social analysis was handed over to the internal auditors for subsequent analysis. Given the outcomes reported from the prior process mining analyses, it would not come as a surprise that for reasons of confidentiality we cannot discuss the outcome of that investigation.

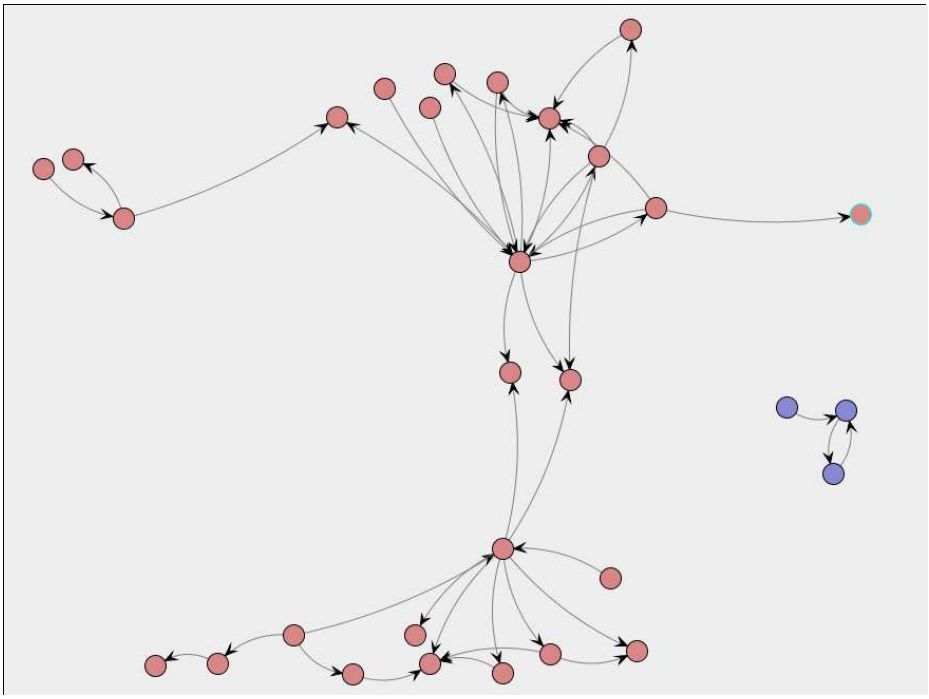

Figure 6: Social Network of the 742 cases without Sign and in violation of controls

7. Conclusion

The purpose of this paper is to establish whether process mining can add value to internal auditing, as has been proposed by Jans et al. (2010). Process mining is the subject of intensive interest by both academics and industry in a variety of domains, but not yet in accounting, and this paper is the first to apply process mining to an actual business in order to assess its efficacy in auditing. In this case study we had the unique advantage of having access to data that had already been audited by the business's internal auditors, thus providing a benchmark to assess the incremental contribution of process mining in uncovering audit relevant information not previously detected by the standard audit procedures.

As a matter of fact, in their review the internal auditors did not find any material ICFR weakness with the procurement process and judged that the business's SAP® controls were

appropriately set to ensure a strong control environment. By contrast, using four different analysis tasks from the process mining toolkit, we were able to identify numerous instances of audit relevant information that warranted further investigation by the internal auditors:

- Three PO's which passed through the procurement process without any Sign or Release, in violation of control procedures.
- 175 violations of the segregation of duty principle that requires Goods Receipt and Release to be undertaken by distinct individuals.
- 265 payments which did not have a matching invoice.
- 3 PO's which did not show a Goods Receipt entry in the system, although the Goods Receipt indicator was flagged.
- 742 cases which did not show a Sign activity, though the conditions for this exception were not met.

We attribute our results to three distinct advantages of process mining over the standard audit procedures used by the internal auditors:

1. The richness of the event log which contains input and meta-data, as well as a comprehensive set of attributes, all systematically ordered by time and originator.
2. The ability to analyze the entire population instead of relying only on a sample.
3. Being able to obtain and visualize a process-view, which also led to more identified ICFR issues in the studies of Kopp and O'Donnell (2005) and Bierstaker et al. (2009).

Of course, the identified ICFR issues represented only a small fraction of the total population, but that by itself does not indicate a reduced value to internal auditing since AU 329 explicitly requires auditors to seek out outliers, they being the most likely indication of fraud and other reporting problems. Moreover, the fact that anomalous transactions are rare demonstrates the power of process mining, particularly considering that the standard audit procedures failed utterly to detect any of these issues.

A limitation of our case study research approach is its lack of generalizability, but that is an issue only for the specific results and not for the general conclusion that process mining can find audit relevant information that standard audit procedures miss. Our sample was chosen essentially randomly and there is no reason to expect that the results would be substantially different with another sample drawn from another process or business. At any rate, the results of this paper strongly argue that research into the application of process mining in the audit

domain continue, both to validate our conclusions and to provide benchmarks to internal auditors for their own process analysis.

References

1. Agrawal, Rakesh, Dimitrios Gunopulos, and Frank Leymann. 1998. Mining process models from workflow logs. In *Advances in Database Technology — EDBT'98*, edited by H.-J. Schek, G. Alonso, F. Saltor and I. Ramos: Springer Berlin / Heidelberg.
2. Alles, M.G., G. Brennan, A. Kogan, M. A.Vasarhelyi. 2006. Continuous Monitoring of Business Process Controls: A Pilot Implementation of a Continuous Auditing System at Siemens. *International Journal of Accounting Information Systems*, Vol.7, 137–161.
3. Bierstaker, J. L., J. E. Hunton, and J. C. Thibodeau. 2009. Do Client-Prepared Internal Control Documentation and Business Process Flowcharts Help or Hinder an Auditor's Ability to Identify Missing Controls? *Auditing: A Journal of Practice & Theory* 28 (1):79-94.
4. Bozkaya, M., J. Gabriels, and J.M. van der Werf. 2009. Process Diagnostics: A Method Based on Process Mining. Paper read at International Conference on Information, Process, and Knowledge Management (eKNOW), February, 1-7, at Cancun, Mexico.
5. Carnaghan, C. 2006. Business process modeling approaches in the context of process level audit risk assessment: An analysis and comparison. *International Journal of Accounting Information Systems* 7 (2):170-204.
6. Cook, J. E., and A. L. Wolf. 1998. Discovering Models of Software Processes from Event-Based Data. *ACM Transactions on Software Engineering and Methodology* 7:215-249.
7. de Medeiros, A. K., A. J. M. M. Weijters, and W. M. P. D. Aalst. 2006. Genetic process mining: A basic approach and its challenges. *Business Process Management Workshops* 3812:203-215.
8. Folino, F., G. Greco, A. Guzzo, and L. Pontieri. 2009. Discovering Multi-perspective Process Models: The Case of Loosely-Structured Processes. In *Enterprise Information Systems-B*, edited by J. Filipe and J. Cordeiro. Berlin: Springer-Verlag Berlin.
9. Greco, Gianluigi, Antonella Guzzo, Luigi Pontieri, and Domenico Saccà. 2006. Discovering Expressive Process Models by Clustering Log Traces. *IEEE Transactions on Knowledge & Data Engineering* 18 (8):1010-1027.

10. Gunther, C. W., and W. M. R. van der Aalst. 2007. Fuzzy mining - Adaptive process simplification based on multi-perspective metrics. In *Business Process Management, Proceedings*, edited by G. Alonso, P. Dadam and M. Rosemann. Berlin: Springer-Verlag Berlin.
11. Jans, Mieke, Michael Alles, and Miklos Vasarhelyi. 2010. Process Mining of Event Logs in Auditing: Opportunities and Challenges. In *International Symposium on Accounting Information Systems*. Orlando.
12. Jans, Mieke, Nadine Lybaert, and Koen Vanhoof. 2010b. Internal fraud risk reduction: Results of a data mining case study. *International Journal of Accounting Information Systems* 11 (1):17-41.
13. Knechel, W.R. 2001. *Auditing Assurance and Risk*. 2nd ed: South-Western College Publishing.
14. Kopp, L. S., and E. O'Donnell. 2005. The influence of a business-process focus on category knowledge and internal control evaluation. *Accounting Organizations and Society* 30 (5):423-434.
15. O'Donnell, Ed, and Joseph J. Schultz Jr. 2003. The Influence of Business-Process-Focused Audit Support Software on Analytical Procedures Judgments. *Auditing: A Journal of Practice & Theory* 22 (2):265-279.
16. Rozinat, A., and W. M. P. van der Aalst. 2008. Conformance checking of processes based on monitoring real behavior. *Information Systems* 33 (1):64-95.
17. van der Aalst, W. 2011. *Business Process Management Center* 2011 [cited February 2011].
18. van der Aalst, W. M. P., H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song, and H. M. W. Verbeek. 2007. Business process mining: An industrial application. *Information Systems* 32 (5):713-732.
19. van der Aalst, W. M. P., M. H. Schonenberg, and M. Song. 2011. Time prediction based on process mining. *Information Systems* 36 (2):450-475.
20. van der Aalst, W. M. P., B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and Ajmm Weijters. 2003. Workflow mining: A survey of issues and approaches. *Data & Knowledge Engineering* 47 (2):237-267.
21. van der Aalst, Wil, Ton Weijters, and Laura Maruster. 2004. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge & Data Engineering* 16 (9):1128-1142.

22. van Dongen, B. F., A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst. 2005. The ProM Framework: A New Era in Process Mining Tool Support. In *Applications and Theory of Petri Nets 2005*, edited by G. Ciardo and P. Darondeau: Springer Berlin / Heidelberg.