A FIELD STUDY ON THE USE OF PROCESS MINING OF EVENT LOGS AS AN ANALYTICAL PROCEDURE IN AUDITING

Mieke Jans

Hasselt University
Hasselt, 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

September 16, 2014*

* We thank seminar participants at the 2011 International Symposium on Accounting Information Systems in Rome, the 2012 Accounting Information Systems and Audit mid-year meeting, the 2012 annual meeting of the American Accounting Association and the 9th International Conference on Enterprise Systems and Logistics for helpful comments. Further feedback is welcome and may be addressed to alles@business.rutgers.edu.

A FIELD STUDY ON THE USE OF PROCESS MINING OF EVENT LOGS AS AN ANALYTICAL PROCEDURE IN AUDITING

Abstract There is a very large research literature examining the use analytical procedures by auditors and proposing either new types of analytical procedures or more effective ways of implementing existing procedures. In this paper we validate the value added that process mining of event logs can provide to auditors when used as an analytical procedure. We apply process mining to procurement data from a leading global bank and demonstrate that it enables us to obtain audit relevant information as defined by SAS 56. In particular, we are able to identify numerous anomalous transactions—including those concerning payments made without approval, violations of segregation of duty controls, and violations of company specific internal procedures—which were not detected by the internal auditors when they conducted their own examination with conventional audit procedures. This paper makes a contribution to the research literature on analytical procedures by verifying the potential of a new methodology not yet used in audit practice; a contribution strengthened by the use of actual data as opposed to the reliance of much of this literature on data obtained from experiments, many with student subjects.

Keywords Process mining, analytical procedures, auditing, event logs.

1. Introduction

Statement of Auditing Standards number 56 (SAS 56) specifies that analytical procedures are an integral part of the audit process and have to be used to test for relationships that exist between transactions.[1] There is a very large research literature examining the use analytical procedures by auditors and proposing either new types of analytical procedures or more effective ways of implementing existing procedures (Hylas and Ashton 1982, Kinney 1987, Loebbecke and Steinbart 1987, Biggs et al. 1988, Knechel 1988, Wright and Ashton 1989, Hirst and Koonce 1996, Wilks 2002, O'Donnell and Schulz 2003, O'Donnell and Schulz 2005, Carpenter 2007, Hunton and Gold 2007, Peecher et al 2010, Brewster 2011).

SAS 56 calls for auditors to use analytical procedures to search for significant differences between the observed data and their expectations models. **Process mining of event logs** is a tool that has been developed over the last decade by computer scientists and statisticians, in collaboration with such leading corporations as SAP and Phillips, to provide a way of understanding how complex business processes operate. Process mining allows businesses to undertake a fact-based identification of problems in their business processes (Van der Aalst, 2011) and hence facilitates the comparison of the actual processes against the designed process. Yet, despite extensive use in such areas as business process improvement, healthcare and network security, and the fact that a major application of process mining is conformance testing—seeing whether business processes in practice match the designed process—it has yet to be implemented in auditing.[2]

The data analyzed by process mining is the **event log** that the auditor constructs from the records maintained by a business's information systems. What is particularly promising about event logs as the basis for analytical procedures is the fact that it consists not only of

[1] *"Analytical procedures are an important part of the audit process and consist of evaluations of financial information made by* ***a study of plausible relationships among both financial and nonfinancial data****. Analytical procedures range from simple comparisons to the use of complex models involving many relationships and elements of data.* ***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"* (AICPA, 1988, emphasis added).

[2] The idea of mining the process in a context of workflow processes was introduced by Agrawal et al. (1998). Over the last decade research in this domain has expanded greatly as many different aspects of business process mining have been developed and investigated by researchers from a variety of disciplines (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). Jans et al (2012) provides a literature review on the process mining literature.

data entered by the auditee, but also meta-data that is recorded automatically and independently of the persons and processes whose behavior is the subject of the audit.[3] This stands starkly in contrast to current analytical procedures which rely almost exclusively on data input by the auditee. Moreover, the potential of process mining goes beyond the greater scope of the data being analyzed, to new forms of analytical procedure tests, which when applied to event log data, may yield insights which exceed those obtainable from commonly used audit tools. In short, both the data being analyzed and the techniques utilized by process mining go well beyond current audit practice and this paper's contribution is that it is the first to actually demonstrate the value added that process mining can provide to auditing when used as an analytical procedure. It thus makes the case that process mining is a valuable addition to the analytical procedure toolkit of auditors and one that warrants follow-up research and implementation in practice.

We demonstrate this value added by applying process mining to actual data obtained from one of the world's largest banks which is located in an advanced Western European nation. Our analysis of this data shows that it is possible to find audit relevant information that was not detected by the bank's own internal auditors when they examined that same data. The fact that the bank has a large and highly professional internal audit group whose integrity has never been questioned further strengthens the argument that process mining gives auditors new information not obtainable otherwise from current audit procedures.

In particular, using process mining of event logs we are able to identify numerous anomalous (unusual) transactions—including those concerning payments made without approval, violations of segregation of duty controls, and violations of company specific internal procedures—which were not detected by the internal auditors when they conducted their own examination with conventional audit procedures, both manual and IT-system based. Whether what we find are indications of fraud, or failed controls, or quite legitimate transactions is a question that only the bank's internal auditors can answer based upon their follow up inquiries. But the outcome of that investigation is secondary as far as the objectives of the paper are concerned, which is to show that when used as an analytical procedure process mining does indeed identify potentially problematic transactions not identified by existing audit technique.

While the generality of the conclusions we can draw from this research is constrained by the fact that we are necessarily restricted to a limited process view of a single business, the

[3] Jans et al (2012)

results support the argument that process mining should be explored as a promising new analytical procedure in auditing. Given the success of this field study, further work by both academic researchers and practitioners is called for in order to better delineate the specific circumstances where process mining can provide the greatest value added.

This paper makes a contribution to the research literature on analytical procedures by verifying the potential of a new methodology not yet used in audit practice; a contribution strengthened by the use of actual data as opposed to the reliance of much of this literature on data obtained from experiments, many with student subjects[A1]. Our paper also opens a new avenue for research into how auditors can best make use of the population-based business process view that process mining promises. As Brewster (2011) states: *Auditing theorists suggest that training auditors to use systems-thinking skills can improve how they learn and use complex entity-level evidence that would normally overwhelm auditors… [This] suggests that new systems-based training methods could help auditors to learn better in the field and develop the skill sets needed to understand complex environments.* Future research should examine whether process mining can help auditors achieve this objective.

The remainder of the paper is organized as follows. The next section develops a protocol for applying process mining as an analytical procedure in auditing. Section 3 presents an overview of the field study. Section 4 then applies to the field site data the protocol for process mining of event logs in auditing developed in section 2. This section demonstrates the value added that process mining can provide to auditors when it is used as an analytical procedure to extract audit relevant information from the event log. Section 5 offers concluding comments. An appendix provides a brief discussion of how to construct an event log.

2. A Protocol for Applying Process Mining as an Analytical Procedure in Auditing

Most businesses of any significant size today store their data digitally thanks to the ubiquity of enterprise resource planning (ERP) systems, such as the SAP™ system used in this field study site. The data recorded by an ERP system includes not only entries made by users of that system (which we refer to as “input-data”) but also information recorded automatically by the system about that input data (which we refer to as “meta-data”), such as the timestamp of transactions and the identity of the person entering the data. That information is located in various tables throughout the ERP system’s data base and to create

an event log that data has to be extracted and assembled into a structured data base that facilitates systematic analysis of the input and meta-data.

For the event log data to be worthy analyzing it has to be assumed that basic IT security controls are in place to preclude the routine user from being able to override or modify automated logging. A~~Of course, a~~ny analytical procedure relies on the integrity of the underlying data and this requirement is not unique to process mining. But none of the bank personnel involved with the purchasing function has super user access to the ERP system which would enable them to override automatic collection of the meta-data and hence we deem that risk as being minimal.

There are four characteristics that need to be extracted from the information system about each event in order to facilitate process mining analysis:

1) The *activity* taking place during the event,

2) The *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.[4]

Critically, at least some of these four characteristics are logged by the ERP systems independent of the originator, meaning that users are required to log into the ERP system with their unique password before being able to enter a transaction and the system timestamps that data entry even if the originator also enters a date themselves.

Once an event log is created, then its analysis~~zing it~~ is essentially a special case of data mining and there are numerous tools available for that purpose. What distinguishes process mining from data mining is that the objective of that analysis is to use data contained in the event log to reconstruct and explore the characteristics of business processes. In addition, process mining is unique in the auditing context in that the fact that the event log contains meta-data allows an independent verification to be made about the behavior of the auditee.

The protocol is developed aiming to obtain the largest amount of audit relevant information at the least possible cost to the auditor. However insightful process mining might prove in auditing, like any analytical procedure it still has to satisfy cost/benefit criteria by

[4] To give a sample entry, the event with unique identifier 01340001 refers to a *Sign* (activity) of *Purchase Order 4603* (process instance) by *Ann Smith* (originator) on *October 5th 2012* (timestamp).

ensuring that it is not too demanding on the resources and time of the audit team. Hence, the protocol aims at focusing systematically on the transactions that warrant further manual investigation.

The steps are the following:

1. **Identify the Designed Process:** this is the starting point of the data analysis since the designed process indicates which activities need to be included in the event log.

2. **Event Log Creation:** At this point in time event log creation is not an automated process since data storage architecture varies across ERP systems. Hence the event log that describes the designed process of interest has to be extracted using deep knowledge about both how the particular ERP system operates, as well as how a given business chooses to handle its data. For example, most IT systems could record more meta-data than they actually do because such recording slows down the system (hence, at an extreme, actual keystrokes could be recorded), and that choice is usually made by the IT department and not by auditors. But many important insights can be obtained even when meta-data is restricted to the minimum that virtually any system records, the *Originator* and *Timestamp* variables.

3. **Process Discovery:** The most important of all insights that process mining can provide is to use the data contained in the event log to show how processes actually operate in a business. The role of business processes has also come increasingly to the forefront of auditing practice since the enactment of the Sarbanes[A2] 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 is to make allowance for the numerous deviations from the designed process model that are ~~inevitably~~ often 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 allowed in ~~built into~~ the ERP system 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. 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 (and if the system lacked sufficient flexibility then line managers will almost inevitably work around the system creating a much larger control and audit risk). Thus, the auditor has to anticipate that there will be deviations from the designed process and will have to use analytical procedures to assess whether such deviations are acceptable, or evidence of control failure. Process mining techniques, when used as an analytical procedure, show not only how processes actually operate, but also allow the auditor to focus on specific discrepancies from the designed process which pose the greatest control risk.[5]

4. **Role Analysis:** while process discovery identifies the business processes, role analysis focuses on the individuals. Having the *Originator* variable in the event log is surely one of the most important sources of value added that process mining offers to auditors since it strips away ~~the cloak of~~ anonymity from the actions of the auditee when inputting data into the ERP system. Thanks to the requirement that users enter their personal login information before having access to the system, the auditor is no longer constrained to have to take the auditee' s word for whom did what when as far as transactions are concerned. Role analysis, as we show, also fits perfectly into the renewed attention that auditors have placed on segregation of duty (SOD) controls in the wake of Section 404.

5. **Attribute Analysis:** The more data the event log informs the auditor about the actual activities being undertaken the more it provides insights into how exactly the discovered process violates the designed process by pinpointing the specific areas where violations are taking place. Instead of simply discovering processes, attribute analysis allows a more subtle examination of when flexibility in the control architecture is appropriate and when it has been abused.

[5] Alternatively, the auditor could put in place systems to monitor and issue alerts when deviations take place from the designed process. In this instance the deviations that were allowed for concerned a margin between the purchase order and the actual invoice of up to 2%, with alerts issued otherwise. But in terms of following order of the activities, the process was quite loosely structured in order to be operationally effective. A major constraint is the effect on the speed of the IT system with such intrusive monitoring.

6. **Social Network Analysis:** The most detailed information that process mining provides auditors arises when role analysis is used to examine anomalous transaction identified using process discovery and attribute analysis, so triangulating in on not just the transaction that warrant manual investigation, but also the players involved with those transactions. This is perhaps the area that warrants the most research into the application of process mining, because it offers the promise—and at this point we would not go further than that—of finally allowing auditors to tackle the most intractable problem that they have always struggled to overcome: collusive fraud.

There is one other critical issue that needs to be addressed: what is the benchmark used to determine the potential value added of process mining? Detecting previously undetected fraud would be strong evidence but that is too high a standard to aim for. Most of what concerns auditors on a day to day basis is seeking evidence of control weaknesses that need rectification in order to achieve a clean Sarbanes Oxley section 404 certification and audit opinion. Moreover, researchers are limited to finding indications of issues that warrant further manual investigation by the bank's own internal auditors. Whether those auditors would share the results of such an investigation with an outside party if fraud is actually found is questionable. It is for these reasons that our benchmark for success in the application of process mining to event logs is the detection of audit relevant information which are transactions that fall within the purview of SAS 56: variations from plausible relationships in the data. That is also the standard adopted in the research literature on analytical procedures. Thus, as Brewster (2011, emphasis added) states: *"When performing analytical procedures, auditors should treat **any** discrepancies between pre-developed expectations and management representations as indicators of heightened misstatement risk."*

3. Field Study Site

The field study is undertaken at a leading European bank which ranks among the top 25 in the world by asset size and which is also subject to the provisions of the Sarbanes Oxley act because of its operations in the United States. The bank's procurement process is the specific focus of the study, chosen because it is a typical, standardized business process in most businesses around the world and hence makes the field study more generalizable. Moreover, procurement represents a large expense item for even this non-manufacturing business, totaling some 1.4 billion Euros in the period covered in this field study and since it is a process with implications for the financial reporting process, it is subject to Section 404 controls under the Sarbanes Oxley act, particularly SOD controls. The greatest hurdle with

any field based research is gaining access to confidential firm data. Few businesses would allow outside researchers to examine very recent transactions and contemporaneous information would not suit our purposes in any case since they would not have been fully audited. But when using data from further back in time two constraints arise: first, the older the data the lower the ability of the internal auditors to effectively investigate any audit relevant information that the process mining protocol uncovers; and secondly, this bank was severely affected by the global credit crisis which began in 2008 and so data from that period lacks generalizability. Given these constraints we were fortunate to gain the cooperation that we did from the bank and we were given permission to extract data from its ERP system up to early 2007. That was sufficiently before the credit crisis began and still within the bank's internal auditors' institutional memory.

Data is available for the construction of event logs for this project because the bank uses SAP™ as its ERP system, including for its procurement cycle. The configuration settings of their ERP system are the focus of internal controls over financial reporting (ICFR) and, the procurement process over the field study time period had already been reviewed by~~received a clean opinion from~~ the bank's internal auditors[A3].

The transactions that were the subject of our field study consist of all the invoices paid during the month of January 2007, which were then traced back to their accompanying PO's. These PO's, with creation dates between 2005 and January 2007, were in turn followed from their start activity *Create PO* to their end activity of *Pay*. If the end activity was not a single payment but rather multiple payments, then the ending activities were cut off after the last payment activity, with this last payment ~~obviously~~ having to take place in January 2007 given our data selection criteria. Hence, the event log only contains completed procurement cycles, a choice which may restrict the types of anomalies that can be detected, but which also prevents false alarms from being generated by POs without a payment yet in the system because the procurement cycle is not yet completed.

Note that we did not take a sample of this procurement data as the bank's own internal auditors did when they reviewed the same processes. Rather, our process mining techniques allow us to analyze the entire population of invoices paid in January 2007, thereby demonstrating the ability of process mining techniques to run on a population set on a monthly basis. Note too, that since we can analyze the entire population of data, the distinction between analytical procedure and tests of detail becomes moot. Indeed, one could perhaps go further and suggest that process mining makes that distinction obsolete as it gives

auditors the ability to dispense with sampling and the use of analysis restricted to aggregated data.

A variety of research tools were used to obtain the necessary details about the procurement process at the business. ~~Guidance~~ Clarification on the way processes operated was obtained from executive officers (both business and information systems specialists), employees at various departments were observed during their job, and the internal user guidelines of the ERP system were consulted. With that comprehensive view of the procurement process in hand the protocol for process mining of event logs in auditing was implemented.

4. Implementing the Process Mining Protocol

4.1 Identify the Designed Process

The flow chart in Figure 1 describes the main activities in the designed procurement process. The process is triggered by the creation of a *Purchase Order* (PO) by an employee. This PO is meant to be *Signed* and *Released* by two distinct and authorized personnel, thereby approving the order for release. Once the *Release* has been taken place, the employee can order the goods from the supplier.[6] 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 and trigger the *Pay* activity.

[6] Since this takes place outside the ERP system the supplier activity is not depicted in Figure 1.

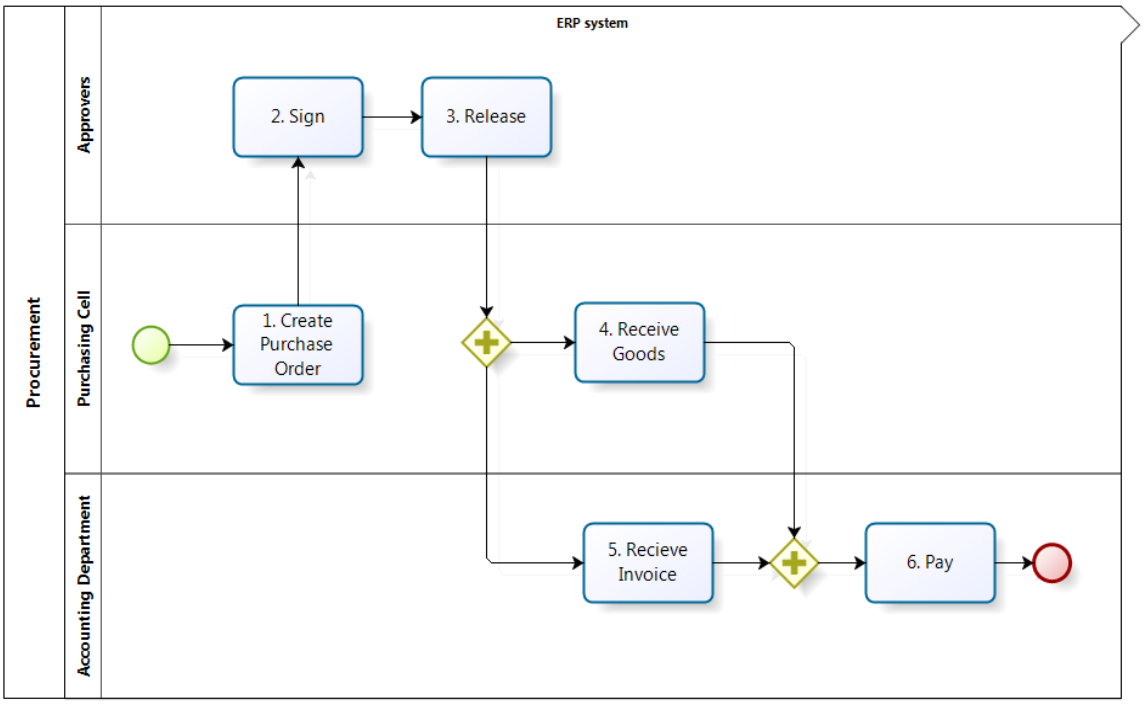

Figure 1: Flow Chart of the Designed Procurement Process

In practice, the procurement process can deviate from the designed model as shown in Figure 1 for operational reasons, as discussed above. For example, changes can be made to the PO after its creation, potentially triggering a new *Sign* and *Release* activity. This would result in an extra activity that is not in the designed 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. This would cause extra activities of *Receive Goods*, perhaps followed by multiple invoices, and hence extra activities of *Receive Invoice*.

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 anomalous outliers that require further investigation as indicated by SAS 56. Hence tests of details are required to supplement the tests of controls. Process mining can aid the auditor in conducting these tests of detail in a more comprehensive and systematic fashion than has been possible up to now, but that necessitates first creating the event log that captures all essential aspects of the designed business process.

4.2 Preliminary Analysis of the Event Log

Full details about the second step of the protocol, the creation of the event log, are relegated to the appendix, since it is secondary to the objective of the paper. Once the event log was available we conducted a preliminary analysis to better understand the way in which purchasing is actually carried out in the bank as opposed to the designed process shown in Figure 1 developed in the first step of the protocol.

The event log consists of 26,185 process instances (i.e. PO lines), involving 181,845 activities and 272 originators. The frequency of activities in the event log is summarized in Table 1. The number of process instances in the 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 and that activity is assigned to each PO line. If the designed 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.

There are less *Sign* activities than PO (lines) created (Table 1), meaning that not every PO is signed, and so nor are the PO lines included in it. On the other hand are there more releases than PO lines. More payments are identified than cases, implying multiple payments on one PO line. Surprisingly~~And, of course, it is striking that~~ almost 60% of POs have changes in them.

Activity	Number	Activity/Create PO
1.Create PO	26185	1.0000
2. Sign	25648	0.9795
3. Release	28748	1.0979
4. GR	24724	0.9442
5. IR	29255	1.1172
6. Pay	31817	1.2151
Change Line	15468	0.5907

Table 1: Frequency of Activities in the Event Log

The average completed procurement process instance consisted of six events, with a minimum of four and a maximum of 390 events. 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.

In short, even this preliminary analysis clearly indicates that there are many issues that need to be examined in detail. The first step in that systematic analysis is discovery analysis.

4.3 Process Discovery

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 designed process model, from which deviations have taken place due to the necessities of operations. 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 meta-data on activities and timestamps. Using traditional analysis techniques would not yield these insights.

The first step in process discovery is to apply a Performance Sequence Analysis to extract all activity patterns that exist in the event log. 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 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. Process mining being such a new audit tool means that there is no benchmark available from procurement and other business processes from a variety of businesses that can serve as a benchmark.[7] Even in the absence of such a basis of comparison most auditors would be surprised that there are such a number and variety of ways in which the procurement process is being carried out given that there is only meant to be a single designed process.

[7] Analysis recently carried out by one of the coauthors of the procurement process at another business revealed well over 1600 distinct patterns. Obviously much remains to be learned about the extent to which processes in practice can deviate from the designed process and the factors that determine the magnitude of those deviations.

		Pattern Frequency		Cumulative total	Paid Value	throughput time (days)			
Pattern	**Sequence**	**#**	**%**	**%**	**Euro**	**avg**	**min**	**max**	**st.dev**
1	Create PO → Sign → Release → GR → IR → Pay	11608	44.3%	44.3%	7 634 968.47	27.78	1	334	20.05
2	Create PO → Change Line → Sign → Release → GR → IR → Pay	6955	26.6%	70.9%	8 178 348.22	32.33	2	343	57.72
3	Create PO → Change Line → Release → IR → Pay	2488	9.5%	80.4%	504 341,51	75.63	3	344	38.99
4	Create PO → Release → IR → Pay	640	2.4%	82.8%	176 540,63	16.8	3	338	26.38
5	Create PO → Change Line → Sign → Release → IR → Pay	491	1.9%	84.7%	1 166 303,16	50.85	6	237	24.07
6	Create PO → Change Line → Sign → Release → IR → GR → Pay	393	1.5%	86.2%	344 583,22	56.36	9	295	40.16

Table 2: Most frequent patterns in the event log

On the other hand 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, but note too that well under half of all process instances correspond to it. Pattern 2 differs from the designed process 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 anomalies were brought to the attention of the bank's internal auditors they stated that there are circumstances (the details of which were not revealed to the researchers) in which it is legitimate to *Release* a PO without a signature. For example, one can hypothesize that some senior managers may have limits below which they can release POs 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 item in question is in reality a service for which there is no act of delivery to a shipping dock (for example, a cleaning service). In fact,

in the case of such services the *Goods Receipt* indicator is supposed to be flagged off in the SAP™ 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-designed patterns in Table 2 warrant some sort of further examination by internal auditors, then what of the remaining 298? On the one hand, one can dismiss them as insignificant, encompassing as they do, less than 14% of all POs. On the other hand, that may be precisely why it is that they should be of the greatest concern to the internal auditors, as SAS 56 indicates with its emphasis on outliers. These 298 involve 40.4 million euros, a non-trivial amount.[A4]

Unfortunately we are unable to pursue this line of inquiry any further in this limited field 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 field 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.

In the absence of a suitable benchmark the discovered process patterns are compared against each other by using a process discovery algorithm to analyze the sequence of activities within the audit trail. Given the large number of patterns, some with a large number of activities (up to a maximum, recall, of 309 activities), visual observation no longer suffices to examine all the audit trails and hence the recourse to a software algorithm to analyze them.

Initially the Fuzzy Miner algorithm of Günther and van der Aalst (2007), is applied 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* 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.

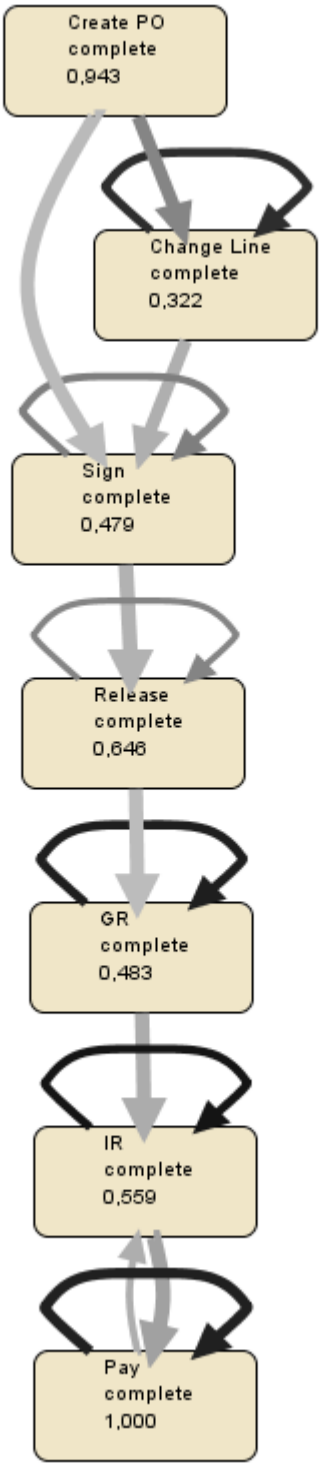

Figure 2. Output of Fuzzy Miner with Default settings to Uncover Core Processes.

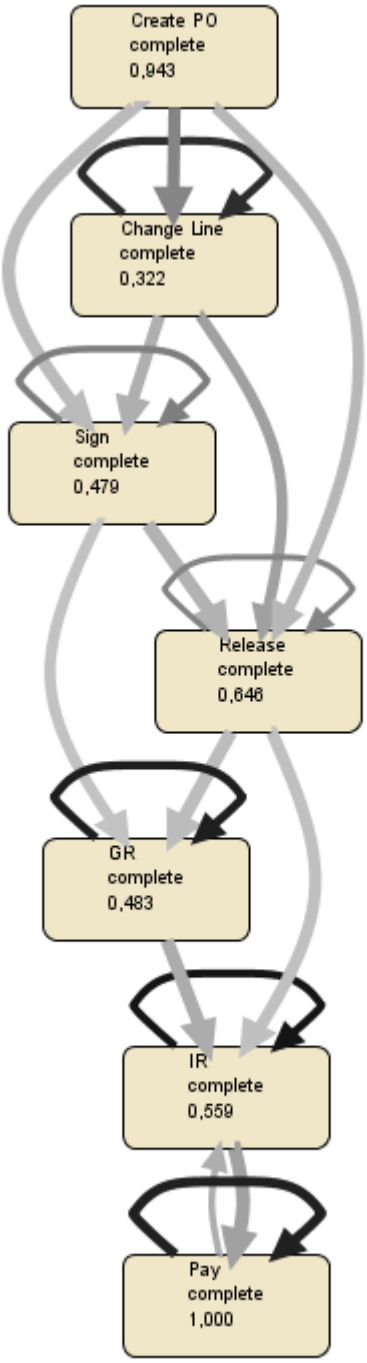

Figure 3. Output of Fuzzy Miner with Lower Threshold Settings to Reveal Less-Frequent Flows

In Figure 3 a more complex process can be observed with extra flows (*edges*). However, it is important to carefully interpret 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, the Linear Temporal Logic algorithm is used to check 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.

	Extra flows	**Occurrences**	**Result**
1	Create PO → GR	0	OK
2	Create PO → Release	739	Verification required for omitting Sign
3	Change Line → Release	2790	Verification required for omitting Sign
4	Sign →GR	11	Further investigation → OK
5	Release → IR	4973	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

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 the bank's 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 attribute 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, patterns 6 and 7 stress the importance of examining whether there exists for each payment a corresponding invoice. This too will be checked in the attribute verification phase. As a final check, it is examined whether each PO process instance has at least one *Release* activity in its pattern. Three cases, out of the population of 26,185 cases, are found 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 this particular event log, but rather, to establish that there is value added to process mining in general. Clearly even these limited process discovery tasks revealed numerous examples of audit relevant information which warrant further investigation by the internal auditors. Next, we undertook other process mining tasks to better define the potential contribution of process mining to auditors.

4.4 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. In order to satisfy Section 404 of the Sarbanes-Oxley Act, businesses have ~~invested heavily on~~been concerned with 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 the bank 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, an Originator-Task matrix was created from the event log which details the number of times an individual executes a particular activity. From this matrix a preliminary check is conducted as to whether some individuals execute an impermissible double role. With 272 originators in this field study , the full table is too large to show, but from the excerpt presented in Table 4 it is found, 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 the population of data.

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 cannot be used to detect all suspect SOD instances and 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 concluded that the first two SOD controls hold without violation in the investigated event log. The fact that an auditor could make such a clear cut assertion as to the efficacy of these two SOD controls for the entire population of transactions across all company personnel is itself indicative of the value added that process mining provides in auditing.

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 this is a further demonstration of the value of role analysis in auditing. This evidence was handed over to the internal auditors for follow up investigation. [A5]

4.5 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 other process mining tools which exploit the information on attributes of the process instances available in the event log.

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 process discovery and role analysis components of the process mining protocol. 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 standalone 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 standalone 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 standalone payments on their account.

These payments were all investigated by the bank's internal auditors 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 payments 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. It was 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 bank's SAP™ system is a shortcoming in the ICFR as 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, but for reasons of confidentiality they could not or would not share with us the outcome of those inquiries.

4.6 Social Network Analysis

The *originator* entry in the event log enables us to construct a social network diagraming all the interactions between all of the 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.

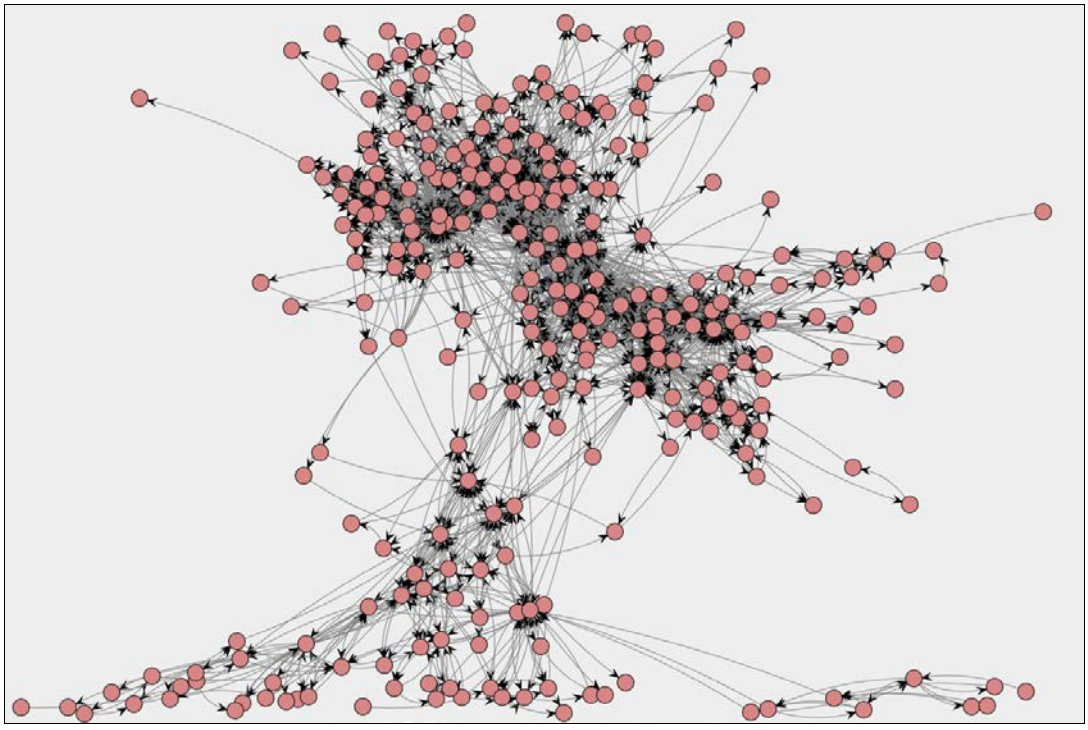

Figure 4: Social Network of all originators in the procurement process

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 was performed. 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.

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. 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[A6].

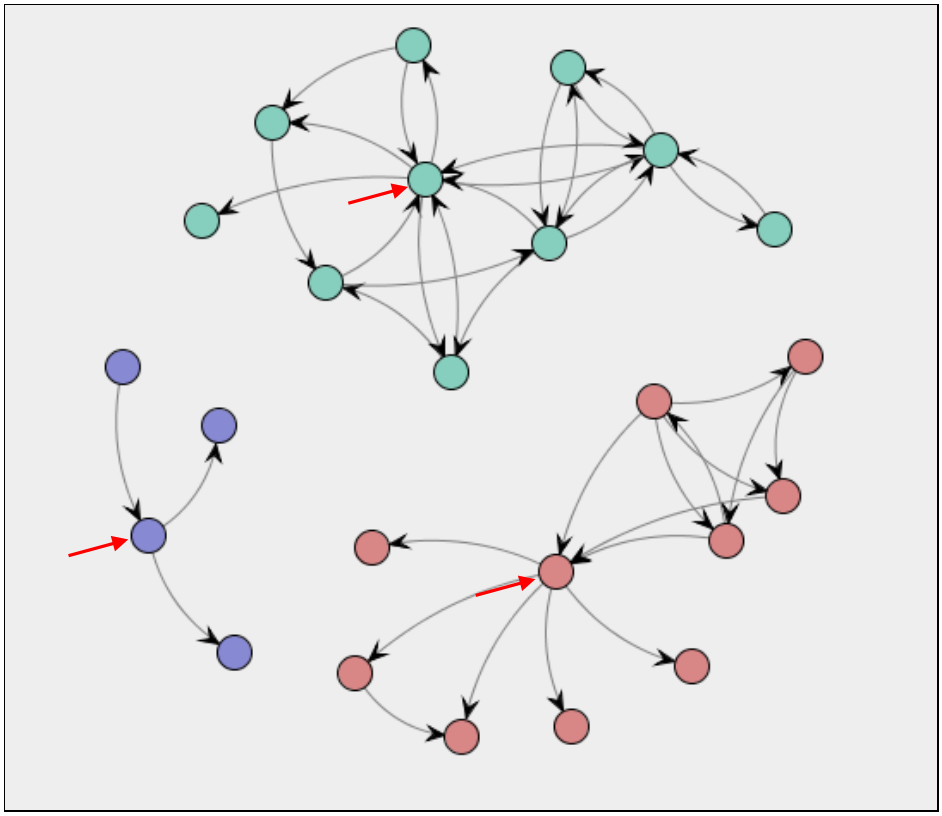

Figure 5: Social Network of 175 cases by three individuals violating SOD controls on Receive and GR

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. Figure 6 shows this social network and indicates 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. 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 arises when, as we have done here, social network analysis is combined with other process mining tasks, such as process discovery and attributes analysis. This exploits the full dimensionality of the data in the event log and focuses attention on the most serious violations of controls[A7].

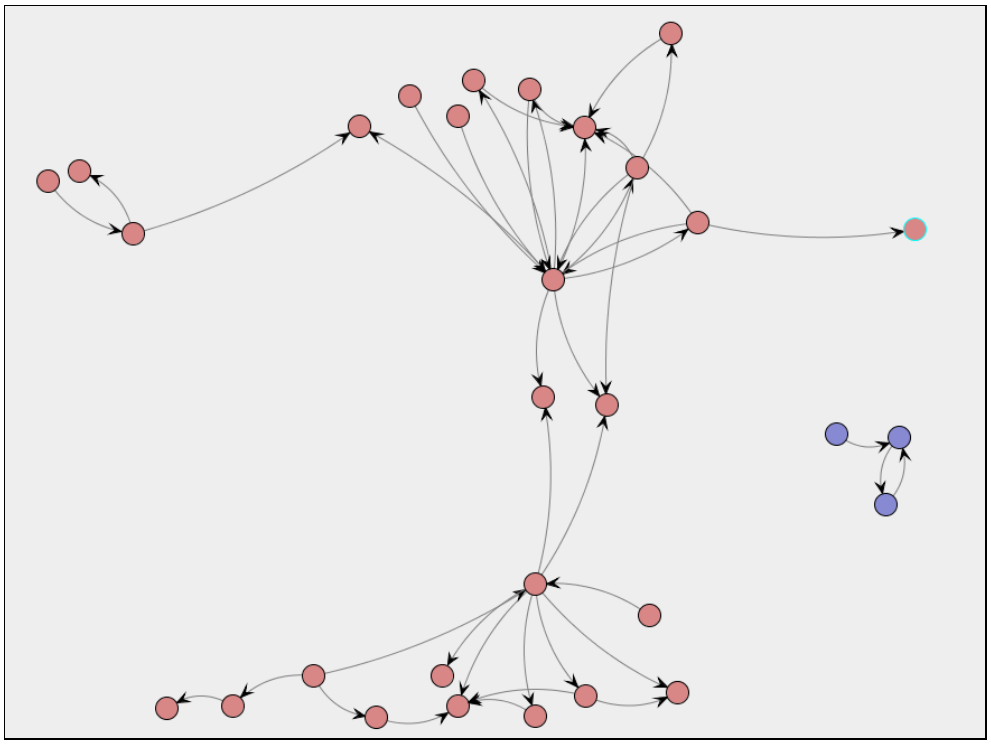

Figure 6: Social Network of the 742 cases without *Sign* and in violation of controls

5. Conclusion

The purpose of this paper was to establish whether process mining can add value to auditors as an addition to their analytical procedures toolkit. This field study had the unique advantage of having access to data 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 standard audit procedures.

In their own review the bank's internal auditors did not find any significant ICFR weakness with the procurement process and judged that its SAP™ controls were appropriately set to ensure a strong control environment. By contrast, the process mining protocol identified numerous instances of audit relevant information that warranted follow-up manual investigation by the internal auditors under SAS 56:

1. Three PO's which passed through the procurement process without any *Sign* or *Release*, in violation of control procedures.
2. 175 violations of the segregation of duty principle that requires *Goods Receipt* and *Release* to be undertaken by distinct individuals.
3. 265 payments which did not have a matching invoice.
4. 3 PO's which did not show a *Goods Receipt* entry in the system, although the *Goods Receipt* indicator was flagged.
5. 742 cases which did not show a *Sign* activity, though the conditions for this exception were not met.

These results can be attributed to two 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 being forced to use only a sample.

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 SAS 56 explicitly requires auditors to seek out outliers as being the most likely indicators of fraud and other control problems. Moreover, the fact that anomalous transactions are rare demonstrates the power of process mining, particularly considering that the standard audit procedures failed to detect any of these issues.

A limitation of field study research is its potential lack of generalizability, but that is an issue only for results specific to the bank and not for the general conclusion that process mining can find audit relevant issues that standard analytical procedures cannot. Since the data analyzed in this study was drawn from a routine process common to all other businesses

there is no reason to imagine that the results on the value added of process mining would be substantially different with a different sample drawn from a different business. The results of this paper strongly argue that further research into the application of process mining in the audit domain needs to be undertaken, 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. Biggs, S., T. Mock, and P Watkins. 1988. Auditors' use of analytical review in audit program design. *The Accounting Review* 63 (January): 148-161
3. 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.
4. Brewster, B.E. 2011. How a systems perspective improves knowledge acquisition and performance in analytical procedures. *The Accounting Review* 86 (3): 915-943.
5. Carpenter, T. D. 2007. Audit team brainstorming, fraud risk identification, and fraud risk assessment: Implications of SAS No. 99. *The Accounting Review* 82 (5): 1119–1140.
6. 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.
7. 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.
8. 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.
9. 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.

10. Hirst, E. and L. Koonce. 1996. Audit Analytical Procedures: A Field Investigation. *Contemporary Accounting Research* Vol 13, No 2, Fall.
11. Hunton, J.E. and A. Gold. 2010. A field experiment comparing the outcomes of three fraud brainstorming procedures: Nominal group, round robin, and open discussion. *The Accounting Review* 85 (3): 911-935.
12. Hylas, R. and R. Ashton. 1982. Audit detection of financial statement errors, *The Accounting Review* Vol. 57 No.4, 751-65.
13. Jans, Mieke, Michael Alles, and Miklos Vasarhelyi. 2012. The Case for Process Mining in Auditing: Sources of Value Added and Areas of Application. *International Journal of Accounting Information Systems* forthcoming.
14. Jans, Mieke, Nadine Lybaert, and Koen Vanhoof. 2010. Internal fraud risk reduction: Results of a data mining field study . *International Journal of Accounting Information Systems* 11 (1):17-41.
15. Kinney, W.R.1987. Attention-Directing Analytical Review Using Accounting Ratios: A Case Study. *Auditing: A Journal of Practice and Theory* Vol. 6, No.2 (Spring), 59-73.
16. Knechel, R. 1988. The effect of statistical analytical review on overall audit effectiveness: A Simulation Analysis. *The Accounting Review* (January): 74-95.
17. Loebbecke, J. and P., Steinbart. 1987. An investigation of the use of preliminary analytical review to provide substantive audit evidence, *Auditing: A Journal of Practice and Theory* Vol. 6 No.2, 74-89.
18. O'Donnell, Ed, and J. 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.
19. O'Donnell, E. and J.J. Schultz, Jr. 2005. The halo effect in business risk audits: Can strategic risk assessment bias auditor judgment about accounting details? *The Accounting Review* 80 (3): 921-939.
20. Peecher, M., M.D. Piercey, J.S. Rich, and R.M. Tubbs. 2010. The effects of a supervisor's active intervention in subordinates' judgments, directional goals, and perceived technical knowledge advantage on audit team judgments. *The Accounting Review* 85 (5): 1763-1786.

21. 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.
22. van der Aalst, W. 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag Berlin Heidelberg.
23. 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.
24. van Dongen, B. F., A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, Agrawal 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.
25. Wilks, T.J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1): 51-71.
26. Wright, A. and R.H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (4): 710-728.

Appendix: A Brief Guide to the Creation of 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?

2) What is the *process instance* that can be linked to the key activities?

These questions are important because they define the content and structure of the event log and their choice is a key judgment call by the auditor. The auditor has to use process analysis to identify the key activities in the designed process. Then it has to be determined what transactions move through these key activities and whether it is possible to collect their data. If the auditor does not have a thorough understanding of the business process, this will reduce the power of the process mining analysis since the discovered model from the event log will fail to include all relevant activities.

The activities selected for the event log are based on the information gathered during the process identification step. The six activities, represented by the six rectangles in Figure 1 above 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), it is available for execution in the SAP™ system as a means of permitting flexibility in the procurement process.

Along with the selection of activities another key decision to be made by the auditor is the choice of the *process instance*. A *process instance* is an object that can be uniquely identified and followed throughout the process by linking its exclusive identifier to the subsequent activities it undergoes. The process instance chosen was the *Purchase Order* line item. The detailed level of a PO item line was preferred 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, the PO lines were followed 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 with signs and releases executed on a PO level as a whole. Thus, these activities are assigned to the PO lines when the activity first takes place, but after a *Change* treat each PO line individually and require that each *Change* itself be signed for and released, and record when that fails to occur.

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 used for the process mining is ProM, which is an established open source framework for executing process mining tasks.[8] In order to apply process mining techniques using the ProM framework the event log needs to be in the MXML format (Mining XML)[9]. 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]

Four characteristics are needed in order to mine a process: an activity, a case, an originator, and a timestamp. Once the first two have been chosen, it was possible 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 who are meta-data, being logged by the SAP™ system automatically and are not based on data input by the originator.

In addition to these four fundamental characteristics there is other information entered by the employee relating to the process instance which 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 that are recorded within the SAP™ system and which is 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 PO item line so that an activity flow, called an *audit trail,* is stored in the event log for each process instance under investigation. For example, the audit trail of the designed process is obviously: *Create PO* → *Sign* → *Release* → *Goods Received* → *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

[8] http://sourceforge.net/projects/prom/.
[9] At the time of the study, only the MXML format was available for structuring an event log. To date, XES is another viable option.
[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. (2010).
[11] In this field study all attributes were input-data, though that needs not be the case in general.

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 case of changing value, the new PO 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.