

# Discovering Organizational Knowledge via Process Mining

Jing Yang<sup>1</sup>[0000-0001-9218-6954]

Queensland University of Technology, Brisbane, Australia  
roy.j.yang@qut.edu.au

**Abstract.** Deploying flexible and proper organizational structures facilitates modern organizations in managing their business processes and the employees involved. To achieve this capability, it requires decision-makers to keep accurate and timely knowledge of their employee groupings. Process mining can help address this need by mining organizational models from event logs, which provide insights on actual resource groupings in the context of business process execution. This PhD research focuses on fulfilling key research gaps in this subfield of process mining, and aims at developing a systematic approach and tool that provide evidence-based support for organizations to understand, evaluate, and improve human resource groupings by using event log data.

**Keywords:** Process Mining · Organizational Model Mining · Event Logs · Conformance Checking

## 1 Introduction

The management of human resources in process execution forms an important perspective in business process management. It is crucial for an organization to have flexible and proper structures around its human resources, especially when facing dynamic demand and fluctuation of employees. A key component in designing organizational structures is to identify the grouping together of employees [3]. Therefore, to deploy effective organizational structures, decision-makers need to maintain accurate and timely understanding of human resource groupings in their organizations. However, such a need can hardly be sufficed by relying on high-level, static organizational charts or the anecdotal knowledge of managers — neither of them offers precise or up-to-date information on resource groupings with respect to the constantly evolving business processes.

In the meantime, analytics performed on employee-related data [4] contributes to deriving detailed and objective insights that connect human resource-related decisions with organizational performance [9]. In particular, event logs extracted from Process-Aware Information Systems [12] can be used as a data source for analyzing human resources in the context of process execution. Event

---

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

logs record the observation of how activities were originated by human resources when executing some process instances at some certain time. Thus, event log data can be utilized for mining knowledge about the behavior and structures of human resources involved in business processes.

Process mining is a discipline that studies knowledge discovery from event log data to facilitate business process improvement. One of its subfields, known as *organizational model mining*, concerns specifically the organizational structures relevant to process execution, thus offers a promising approach to the analytics of resource groupings using event log data. However, research on organizational model mining remains largely under-explored, compared to many other topics in process mining. As evidenced in our literature review, several research gaps remain open and limit the insights on human resource groupings gained from mining event logs.

To this end, the proposed PhD project focuses on the topic of organizational model mining. We aim to address the key research gaps by developing a systematic approach consisting of methods and software tools to provide evidence-based support for organizations to understand, evaluate and improve their human resource groupings based on event log data.

The remainder of this report is structured as follows. Section 2 reviews the literature. Section 3 elaborates on the research questions. Section 4 discusses the research design. Section 5 summarizes the current progress. Finally, Section 6 outlines the expected contributions of this research.

## 2 Literature Review

Organizational model mining is the focus of this research and one of the four topics of process mining from the organizational perspective. To analyze the literature, we employed the classic view of discovery, conformance, and enhancement. The following shows the results.

*Discovery.* This is the most common topic addressed in the existing organizational model mining research. Given an event log as input, an organizational model is constructed to describe the actual resource groupings in the related process. Our literature analysis shows three issues in terms of model discovery.

A typical input event log often records information on multiple dimensions of process execution, i.e., activity, case, and time. However, the majority of the extant discovery methods (e.g., [6,2]) consider merely the activity dimension, for example, the frequency of resources executing similar activities or the handover of work between resources executing consecutive activities. On the other hand, information related to case and time is rarely explored. Only [11] exploits log information on resources participating in the same cases for discovering organizational models. As a result, organizational models in the literature are unable to capture resource groupings that follow patterns on the case and time dimension, for example, project teams and shift workers.

Uncovering the groupings of resources is addressed by most of the literature. However, only a few existing works (e.g., [11]) describe how the discovered

resource groups were involved in process execution. The lack of description of resource groups poses a challenge to understanding the behavior and performance of groups based on discovered organizational models.

The last issue concerns the evaluation of discovery outputs, for which there are three strategies in the literature. The first one compares the discovery results with domain knowledge, such as official organizational structures [6,2]. Clearly, this relies on the availability of prior knowledge, while risks being flawed as there is no guarantee whether the reality has already deviated from the referenced information. The second strategy is assessing the effectiveness of the techniques applied for model discovery. Such evaluations depend on specific techniques, and hence organizational models discovered using different techniques cannot be compared on the same basis. The third strategy considers evaluation by validating the feasibility of proposed methods through experiments on synthetic or real-life event logs, e.g., [5]. So far, none of the existing studies has explicitly considered using input event logs as references for evaluating the discovery results.

*Conformance.* For conformance checking, both an event log and an organizational model are required as inputs, and the outputs are the commonalities or discrepancies found after comparing the log data with the model. Checking the conformance of organizational models can support the evaluation of human resource groupings by exploring their behavior using event logs.

Based on reviewing the literature, it is found that no existing research on organizational model mining directly addresses the issue of conformance checking. Instead, the most relevant topics are rule mining (e.g., [10]) and multi-perspective conformance checking (e.g., [8]). However, it is worth mentioning that neither concerns group-level issues but rather focuses on individual resources.

*Enhancement.* The enhancement issue refers to exploiting event log data for extending or improving existing models. Enhancing organizational models can provide actionable knowledge to decision-making in terms of how to apply the insights from discovery and conformance checking to improve resource groupings.

Some literature has studied how to extend organizational models using event logs. Examples include enriching a set of resource groups with interaction information [11] in order to reveal communication patterns among groups; and utilizing time information to track the changes of organizational models over a certain period of time [1]. On the other hand, some have researched how to improve existing organizational models through simulation, e.g., in [7] event log data is used for analyzing and reducing the communication costs among resources in process execution.

*Conclusions.* Several open research gaps are found by analyzing the state-of-the-art. For one, when the major research focus is on the discovery of organizational models, there remain key issues in the extant discovery methods: (1) They lack in consideration for the multiple dimensions of business processes, and (2) most of the discovered models neglect the description of how discovered resource groups are involved in executing processes, and (3) an appropriate evaluation method

that compares a discovered model against the input event log is still missing. On the other hand, the topics of conformance checking and enhancement in organizational model mining remain largely under-explored.

### 3 Research Problem

Research gaps revealed in the literature analysis lead to several interesting research questions (RQ) being studied in this research.

***RQ1.** How to model and discover resource groupings based on event logs?*

The starting point is to discover organizational models that characterize the actual resource groupings using event logs. A comprehensive discovery approach is needed, which should consider the multiple dimensions of business processes and should construct models that properly characterize resource groups in process execution. Furthermore, the discovery results need to be evaluated appropriately against the input event logs.

***RQ2.** What are the possible aspects and criteria for analyzing resource groupings in process execution?*

***RQ3.** How to analyze resource groupings based on the established aspects and criteria?*

Analyzing resource groupings based on event logs is the prerequisite for improving them. The lack of state-of-the-art research on the conformance checking and enhancement issues in organizational model mining motivates the study of these questions. To address RQ2 and RQ3, it requires formalizing a set of dimensions and the corresponding methods or measures for assessing organizational models. Potential ideas include (1) their conformance to the reality as reflected by event logs, and (2) their appropriateness with regard to organizational design principles in management science. It is also worth “diagnosing” resource groupings using event logs, for example, to detect performance issues related to groups and member resources, and how these issues impact the entire process.

***RQ4.** How to enhance resource groupings in business processes in order to empower human resources and improve business process performance?*

The final research question concerns how to utilize the outcomes from discovering and analyzing resource groupings to improve existing groupings and thus their relevant process performance. To address this question, methods need to be developed, which can produce alternative models that resolve issues uncovered from analyses or inform redesigns of groupings to fulfill improvement targets. For example, a new organizational model can be derived based on revising the “as-is” one, which reallocates the groupings of employees to achieve better designation of responsibilities according to their frequent behavior and performance.

## 4 Methodology and Design

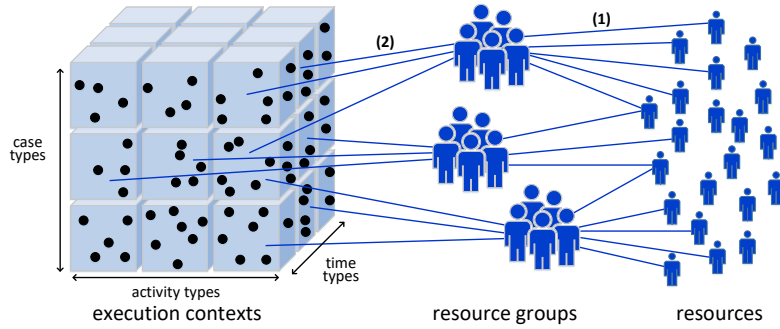
Design Science Research (DSR) methodology is applied when conducting this research. Guided by the iterative process of DSR, a research design was created which outlines the tasks to be addressed as follows.

1. *Identifying Problems*: This has been addressed by conducting a literature review, with a focus on the topic of organizational model mining as well as other relevant work of process mining from the organizational perspective. Topics in other related fields were also examined, including human resource analytics in management, data mining and data visualization techniques in computer science.
2. *Designing Methods*: Methods and software tools will be designed and developed. Knowledge will be drawn from existing research or software related to process mining and human resource analytics for designing. The development of software tools will be achieved following software engineering principles and practices, and be carried out in alignment with some of the open-source process mining software offerings.
3. *Demonstration and Evaluation*: Experimentation is used for testing the designed and developed methods and tools, which is done by (1) selecting and preparing event log datasets for experiments, and then (2) designing and conducting experiments using selected event log datasets to test the proposals. Conclusions from the experiments will be used to refine the design and development of methods and tools. Furthermore, case studies are also planned for evaluating the applicability of the developed methods and tools in real organizations.
4. *Communicating Research*: The research design and outcomes will be communicated to the research community in venues such as doctoral consortiums, conferences, and journal publications. Also, this research will be communicated with HR and workforce planning practitioners and analysts in real organizations, which will be carried out as part of the planned case studies.

## 5 Current Progress

This section summarizes the current stage of the reported research.

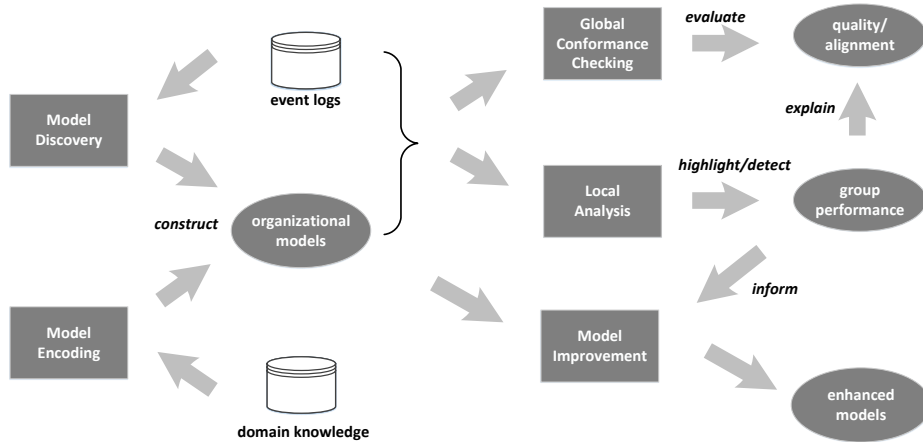
*Novel Definition of Organizational Models*. In view of the current issues of organizational models discovered from event logs, a novel model definition is first proposed. Compared to the literature, this novel definition specifies not only the clustering of individual resources into groups, but also the representation of resource groups' behavior in process execution. This is achieved by introducing the concept of *execution contexts*. Execution contexts are based on viewing resource actions in process execution through the lens of certain meaningful classifications of activities, cases, and times, and therefore turning event logs into data samples of resource groups and their behavior. Using execution contexts enables



**Fig. 1.** Illustration of the novel definition of organizational models [13]

an organizational model to capture various ways of how employees are grouped, such as shift workers or employee teams dedicated to specific customers. Fig. 1 shows an illustration of the novel organizational models. The next step of research considers two tasks. For one, more process dimensions will be included in execution contexts given event log datasets with additional information, e.g., locations of human resources. For another, it is also worth exploring a semi-automatic approach to deriving execution contexts from event logs. Currently, they are derived by manually designating the classifications of activities, cases, and times based on prior knowledge of event logs.

*Conceptual Framework.* Built upon the novel model definition, a conceptual framework has been established to outline the critical research components to address the void of literature on organizational model mining. Fig. 2 shows the framework and the remainder of this section explains each of these components.



**Fig. 2.** A conceptual framework for organizational model mining in this research

*Organizational Model Discovery.* Organizational models can be constructed from either (1) applying a model discovery approach using event logs or (2) applying a model encoding approach using domain knowledge of managers or other official documentation about organizational structures. The former captures the actual groupings of resources in process execution, while the latter captures the official or de jure groupings of resources. Up to date, an initial approach has been developed to discovering models from event logs. Future work considers improving this model discovery approach and developing a model encoding approach.

*Global Conformance Checking Measures.* Global conformance checking is to evaluate the quality of discovered models against event logs or the alignment between official models and the reality. Motivated by the research on process model discovery, the measures of *fitness* and *precision* have been proposed for calculating how well an organizational model represents the actual groupings of resources and their behavior in the reality. The next step of research concerns validating these proposed measures.

*Local Analysis Measures.* Unlike global conformance checking, local analysis is conducted on the level of resource groups or group members instead of the entire organizational model. This part of the research aims at providing methods for highlighting performance insights or diagnose potential issues related to resource groupings. Moreover, local analysis can facilitate explaining the results of global conformance checking performed on organizational models. Currently, an initial set of analysis measures have been established by drawing insights from selected management literature on performance measures of employee teams and departments. A corresponding data visualization tool has been developed to assist in performing local analysis. Based on these, a systematic literature review will be conducted to extend the set of measures for more comprehensive local analyses.

*Tool Development.* An open source software tool was designed following the conceptual framework to demonstrate the ideas proposed in this research, assist in experimentation, and help communicate the outcomes to the research community as well as to practitioners in real organizations. The tool<sup>1</sup> is under active development alongside the progress of this research.

## 6 Contributions

The contributions of this PhD research are expected to be two-fold. From a research point of view, this research will contribute to the field process mining by extending the research on mining organizational models. It will also contribute to the studies on human resource analytics by introducing event log data as an accessible data source for analyzing employee groups, and providing the framework and methods for doing so. As such, this research is expected to promote the value of process mining techniques in the field of human resource management.

<sup>1</sup> Software tool being developed in this project: <https://royjy.me/to/orgminer>

From a pragmatic point of view, the research outcomes can empower managerial teams in organizations by offering them a repeatable and systematic means to gain actionable insights on employee groupings from event log data, and therefore support them in making more guided decisions. Thus, employees will also benefit from the results of improved decision-making by, e.g., engaging in more suitable groups and having more tailored assignments of work.

## Acknowledgments

This research is supervised by Dr. Chun Ouyang, Prof. Arthur ter Hofstede and Prof. Wil van der Aalst, and supported by an Australian Government Research Training Program (RTP) Scholarship.

## References

1. Appice, A.: Towards mining the organizational structure of a dynamic event scenario. *J. Intell. Inf. Syst.* **50**(1), 165–193 (2018)
2. Burattin, A., Sperduti, A., Veluscek, M.: Business models enhancement through discovery of roles. In: *IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. pp. 103–110 (2013)
3. Daft, R.L., Murphy, J., Willmott, H.: *Organization theory and design*. Cengage learning EMEA (2010)
4. Davenport, T.H., Harris, J., Shapiro, J.: Competing on talent analytics. *Harvard business review* **88**(10), 52–58 (2010)
5. Hanachi, C., Gaaloul, W., Mondri, R.: Performative-Based Mining of Workflow Organizational Structures. In: *Proceedings of the 13th International Conference on E-Commerce and Web Technologies*. pp. 63–75. Springer (2012)
6. Jin, T., Wang, J., Wen, L.: Organizational modeling from event logs. In: *International Conference on Grid and Cooperative Computing (GCC)*. pp. 670–675 (2007)
7. Lee, J., Sung, S., Song, M., Choi, I.: A business process simulation framework incorporating the effects of organizational structure. *International Journal of Industrial Engineering* **22**(4), 454–466 (2015)
8. de Leoni, M., van der Aalst, W.M.P., van Dongen, B.F.: Data- and Resource-Aware Conformance Checking of Business Processes. In: Abramowicz, W., Kriksciuniene, D., Sakalauskas, V. (eds.) *Proceedings of the 15th International Conference on Business Information Systems (BIS 2012)*. pp. 48–59. Springer (2012)
9. Marler, J.H., Boudreau, J.W.: An evidence-based review of HR Analytics. *The International Journal of Human Resource Management* **28**(1), 3–26 (2017)
10. Schönig, S., Cabanillas, C., Jablonski, S., Mendling, J.: A framework for efficiently mining the organisational perspective of business processes. *Decis. Support Syst.* **89**, 87 – 97 (2016)
11. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. *Decis. Support Syst.* **46**(1), 300 – 317 (2008)
12. van der Aalst, W.M.P.: *Process Mining: Data Science in Action (2nd Ed.)*. Springer, Berlin, Heidelberg (2016)
13. Yang, J., Ouyang, C., van der Aalst, W.M.P., ter Hofstede, A.H.M., Yu, Y.: *OrgMining 2.0: A novel framework for organizational model mining from event logs* (2020), <https://arxiv.org/abs/2011.12445>