Graph Learning for Process Discovery

TESI DI LAUREA MAGISTRALE IN

COMPUTER SCIENCE AND ANGINEERING

|  |  |
| --- | --- |
| Student ID:  Advisor:  Co-advisor:  Academic Year: | 010499057  Matteo Matteucci  Pietro Portolani  202122 |

Author: **Matteo Badenchini**

# Abstract

Here goes the Abstract in English of your thesis followed by a list of keywords. The Abstract is a concise summary of the content of the thesis (single page of text) and a guide to the most important contributions included in your thesis. The Abstract is the very last thing you write. It should be a self-contained text and should be clear to someone who hasn’t (yet) read the whole manuscript. The Abstract should contain the answers to the main scientific questions that have been addressed in your thesis. It needs to summarize the adopted motivations and the adopted methodological approach as well as the findings of your work and their relevance and impact. The Abstract is the part appearing in the record of your thesis inside POLITesi, the Digital Archive of PhD and Master Theses (Laurea Magistrale) of Politecnico di Milano. The Abstract will be followed by a list of four to six keywords. Keywords are a tool to help indexers and search engines to find relevant documents. To be relevant and effective, keywords must be chosen carefully. They should represent the content of your work and be specific to your field or sub-field. Keywords may be a single word or two to four words.

**Key-words:** here, the keywords, of your thesis.

# Abstract in lingua italiana

Qui va l’Abstract in lingua italiana della tesi seguito dalla lista di parole chiave.

**Parole chiave:** qui, le parole chiave, in italiano, della tesi.

# Contents

[Abstract i](#_Toc110788350)

[Abstract in lingua italiana iii](#_Toc110788351)

[Contents v](#_Toc110788352)

[SECTION I – Process Mining 7](#_Toc110788353)

[1. Process Mining 8](#_Toc110788354)

[1.1 Overview and objectives 8](#_Toc110788355)

[1.1.1 Play-in 9](#_Toc110788356)

[1.1.2 Play-out 9](#_Toc110788357)

[1.1.3 Replay 9](#_Toc110788358)

[1.2 Data 9](#_Toc110788359)

[1.3 Models 9](#_Toc110788360)

[1.3.1 Petri Nets and Workflow Nets 9](#_Toc110788361)

[1.3.2 Process Trees 9](#_Toc110788362)

[2. Process Discovery 10](#_Toc110788363)

[2.1 Overview and Objectives 10](#_Toc110788364)

[2.2 State of the Art 10](#_Toc110788365)

[2.2.1 Principal Miners 10](#_Toc110788366)

[2.2.2 The work of Sommers 10](#_Toc110788367)

[3. Bibliography 11](#_Toc110788368)

SECTION I – Process Mining

# Process Mining

## Overview and objectives

Process mining can be defined as the bridge between data science and process science[[1]](#footnote-1) (Aalst, 2016). The goal of process mining is to discover, analyze and improve existing processes by extracting the relevant information from event data stored inside information systems.

Process mining represents the core of the so-called Business Process Management (BPM) lifecycle, which describes the distinct phases of managing a particular business process.

(re)design

model

optimize

monitor

execute

After being designed in the *design* phase, the model is transformed into a running system in the *model* phase. Once an implementation of the model is available, it is executed, and the behavior analyzed. In the *monitor* and *optimize* phases, the data generated by the execution is analyzed and compared with the desired behavior of the blueprint. If necessary, the model is redesigned, in the attempt to optimize (or remove) bottlenecks, include observed behavior and remove unobserved one.

The *monitor* and *optimize* phases are especially important, since they allow to close the cycle, and therefore to incorporate into the design information about the actual execution of the process in a systematic way.

In the filed of process mining, we can identify three different branches (or type of tasks) that conduct the distinct phases of the BPM lifecycle: play-out, play-in and replay.

### Play-in

### Play-out

### Replay

## Data

As mentioned above, the starting point of any process mining task is event data, which is usually stored inside information systems in the form of logs.

## Models

### Petri Nets and Workflow Nets

### Process Trees

# Process Discovery

## 2.1 Overview and Objectives

## 2.2 State of the Art

### 2.2.1 Principal Miners

#### 2.2.1.1 Alpha

#### 2.2.1.2 Heuristics

#### 2.2.1.3 Inductive

#### 2.2.1.4 ILP

### 2.2.2 The work of Sommers

# Bibliography

Aalst, W. v. (2016). *Process MIning: Data Science in Action.* Springer.

1. The discipline that focuses on process modeling and the implementation of processes (not necessarily in a data driven way). [↑](#footnote-ref-1)