

Knowledge process mining and optimization



Tadej Štajner
Carolina Fortuna
Dunja Mladenić
Marko Grobelnik

SIKDD 2010
12 October 2010
Ljubljana, Slovenia

Outline

- ▶ Motivation
- ▶ Models
 - ▶ Context, Action, Process
- ▶ TaskMiner
- ▶ Conclusions and future work

Motivation

- ▶ A knowledge worker performing tasks accross projects
 - ▶ A worker whose activity mainly revolves around accessing and producing knowledge resources
 - ▶ In our case, the work consists of reading and writing e-mail, browsing web pages, reading and authoring documents
- ▶ How to represent and understand the underlying process?
 - ▶ ... to help the process analyst optimize the process

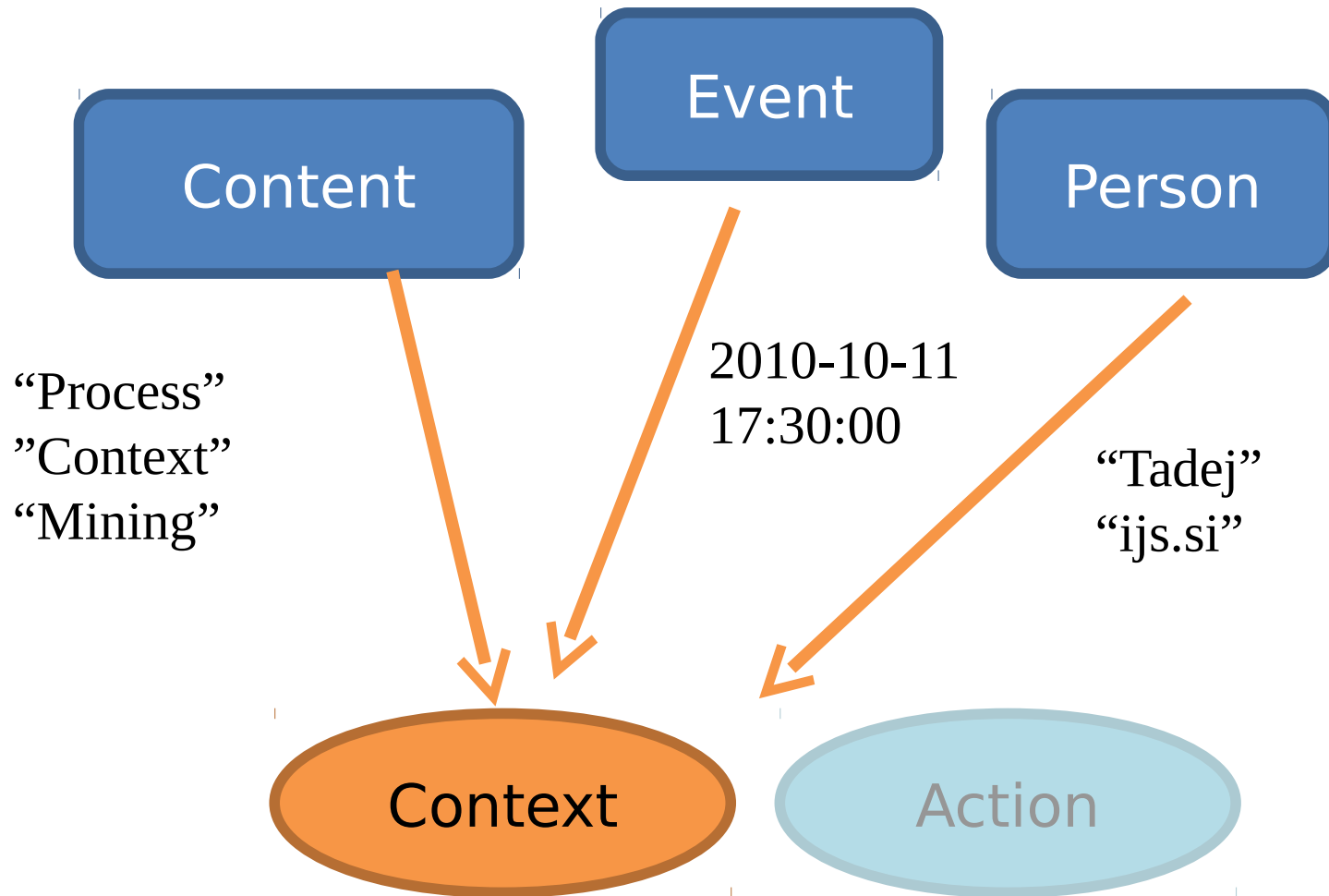
Model

- ▶ A knowledge worker operates in different **contexts**, where each context is an instantiation of a **process**, which is described by **action** patterns
- ▶ Requires solving the following:
 - ▶ What are the contexts?
 - ▶ What are the actions?
 - ▶ How to represent a process that it:
 - ▶ allows for prediction
 - ▶ is interpretable

Context

- ▶ **A grouping of information resources for a particular need**
- ▶ Knowledge workers usually work in several different contexts
- ▶ At any point in time, a knowledge worker is working in some context, eventually switches to another
- ▶ **How to describe/distinguish contexts?**

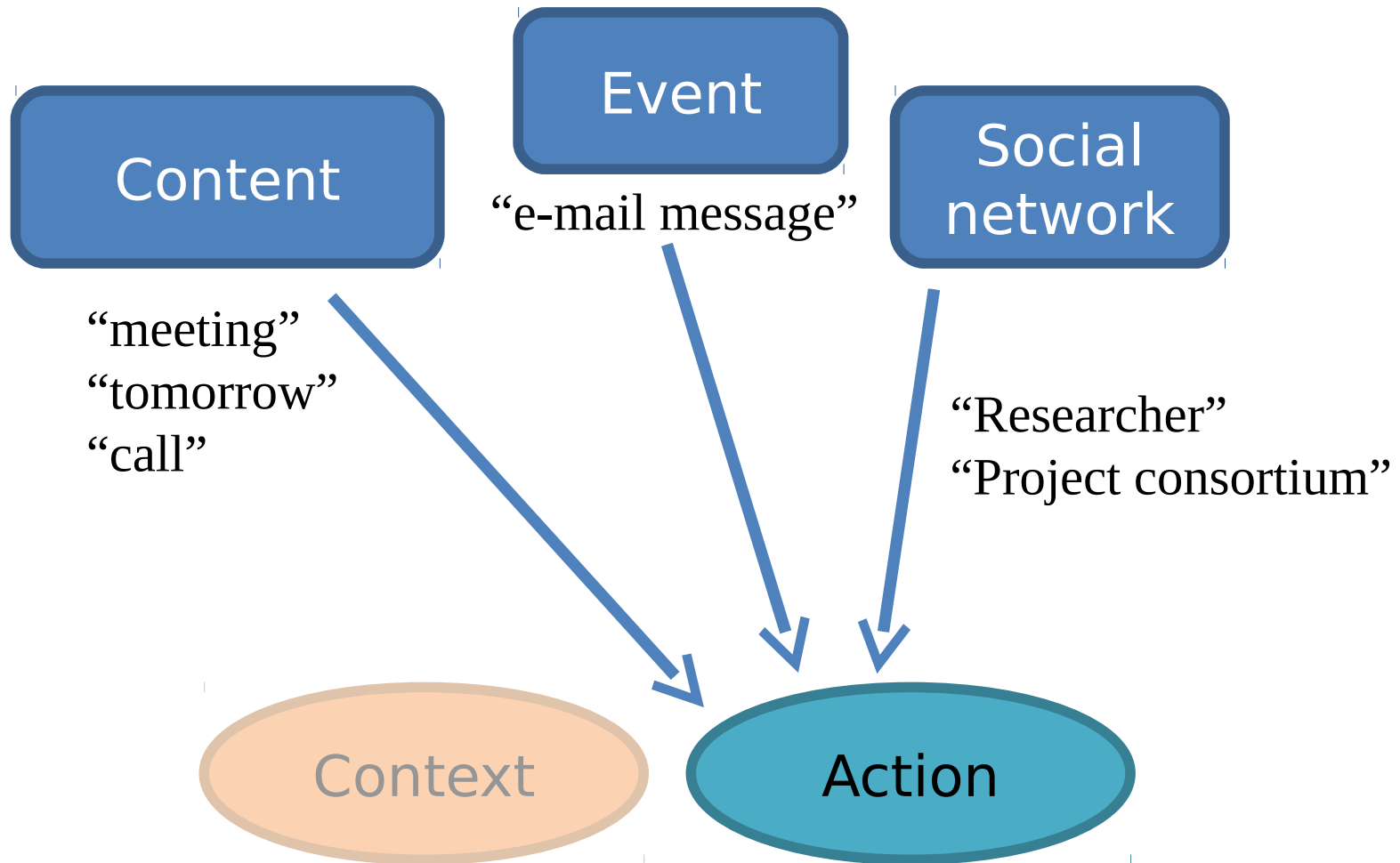
Context model features



Action

- ▶ **An atomic step in the execution of a process**
 - ▶ Context independent
 - ▶ Should represent some function, appropriate for the domain
 - ▶ Each event is a manifestation of some action
-
- ▶ **How to describe/distinguish actions?**

Action model features



Context and action mining

- ▶ Each event is executed in a context
 - ▶ *I have sent an invoice to **ACME Inc.***
- ▶ Each event is a manifestation of an action
 - ▶ *I have **sent an invoice** to ACME Inc.*
- ▶ Both mining tasks are similar – both context and action models can be constructed as clusterings of events using different feature sets
 - ▶ Clustering on context features gives us context models
 - ▶ Clustering on action features gives us action models

Process

- ▶ **A dynamic model of action-to-action transitions**
- ▶ Purposes:
 - ▶ Understanding the real process (not just prescribed)
 - ▶ Predicting user behaviour
- ▶ Possible formalizations are Petri nets, EPC, Markov Models
- ▶ Knowledge worker logs tend to have relatively primitive events, which result in complex models
 - ▶ Solution: apply a measure of statistical significance on the observed patterns

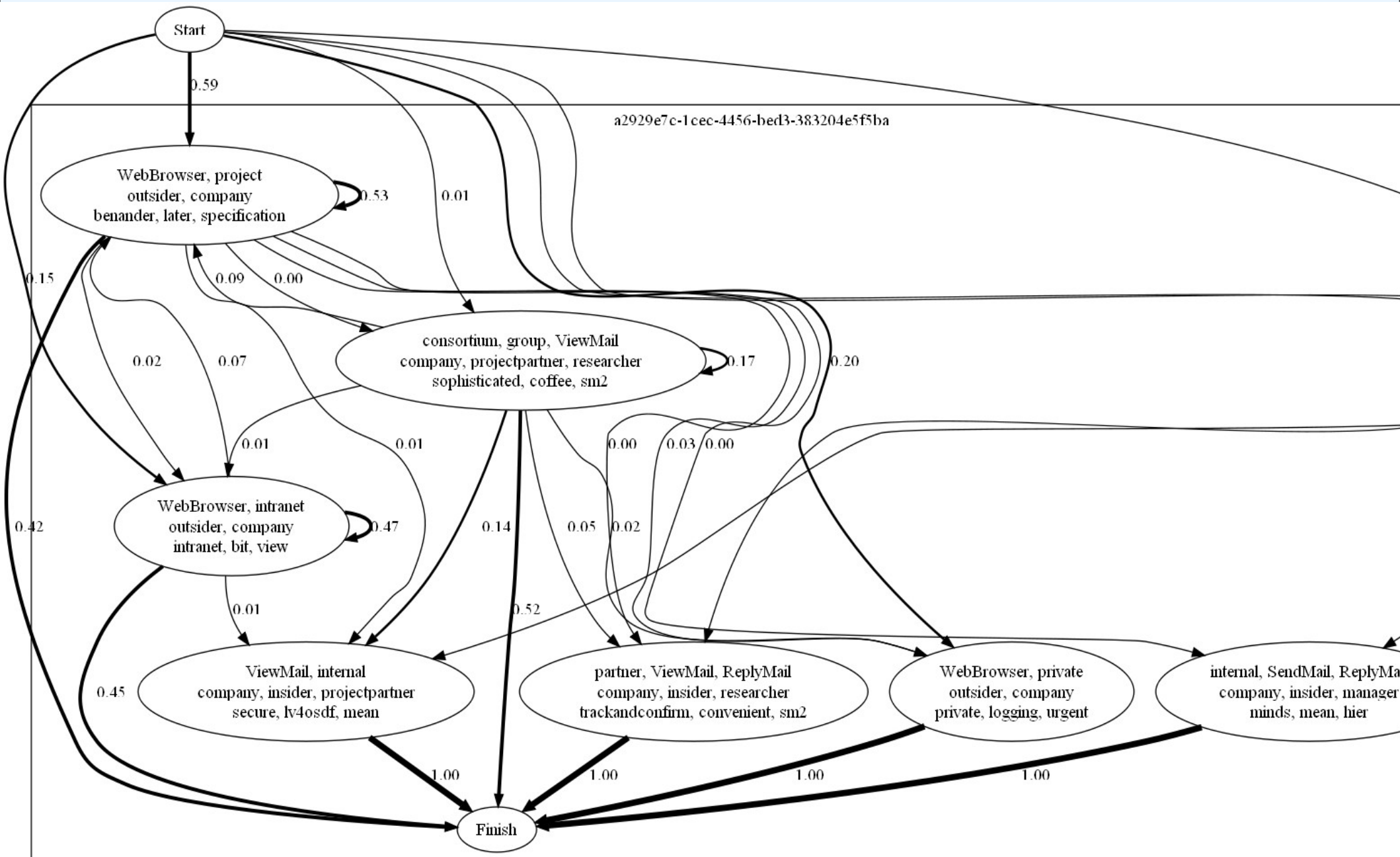
Process mining

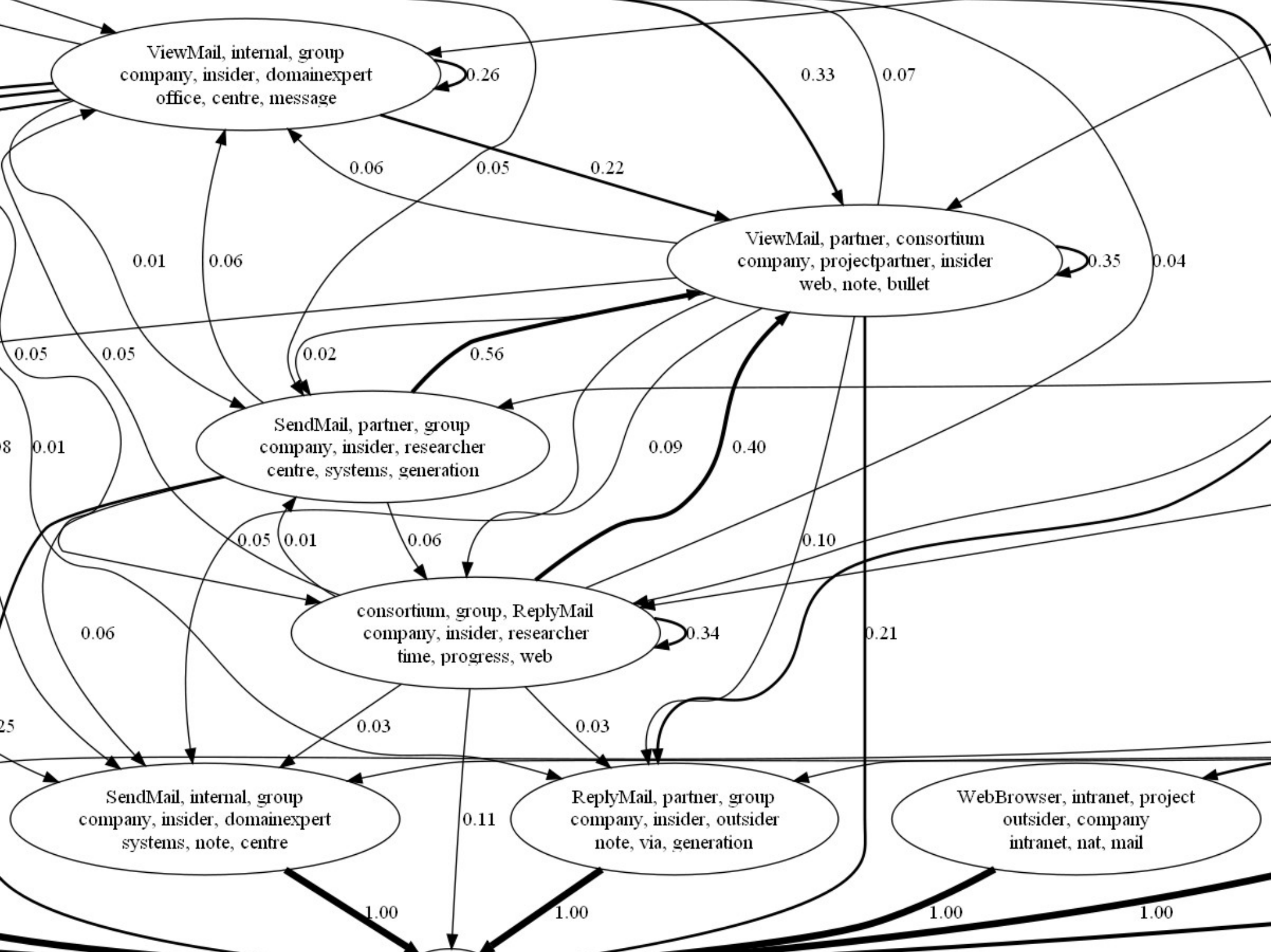
- ▶ **Step 1:** discover the contexts and actions
- ▶ **Step 2:** partition the events (e-mail messages, document accesses..) by context
- ▶ **Step 3:** for each context, take all events and the actions they represent and construct a process model (probabilistic deterministic finite automaton) using transitions from one action to another
- ▶ **Step 4:** prune the obtained PDFA to contain only statistically significant transitions

Process optimization

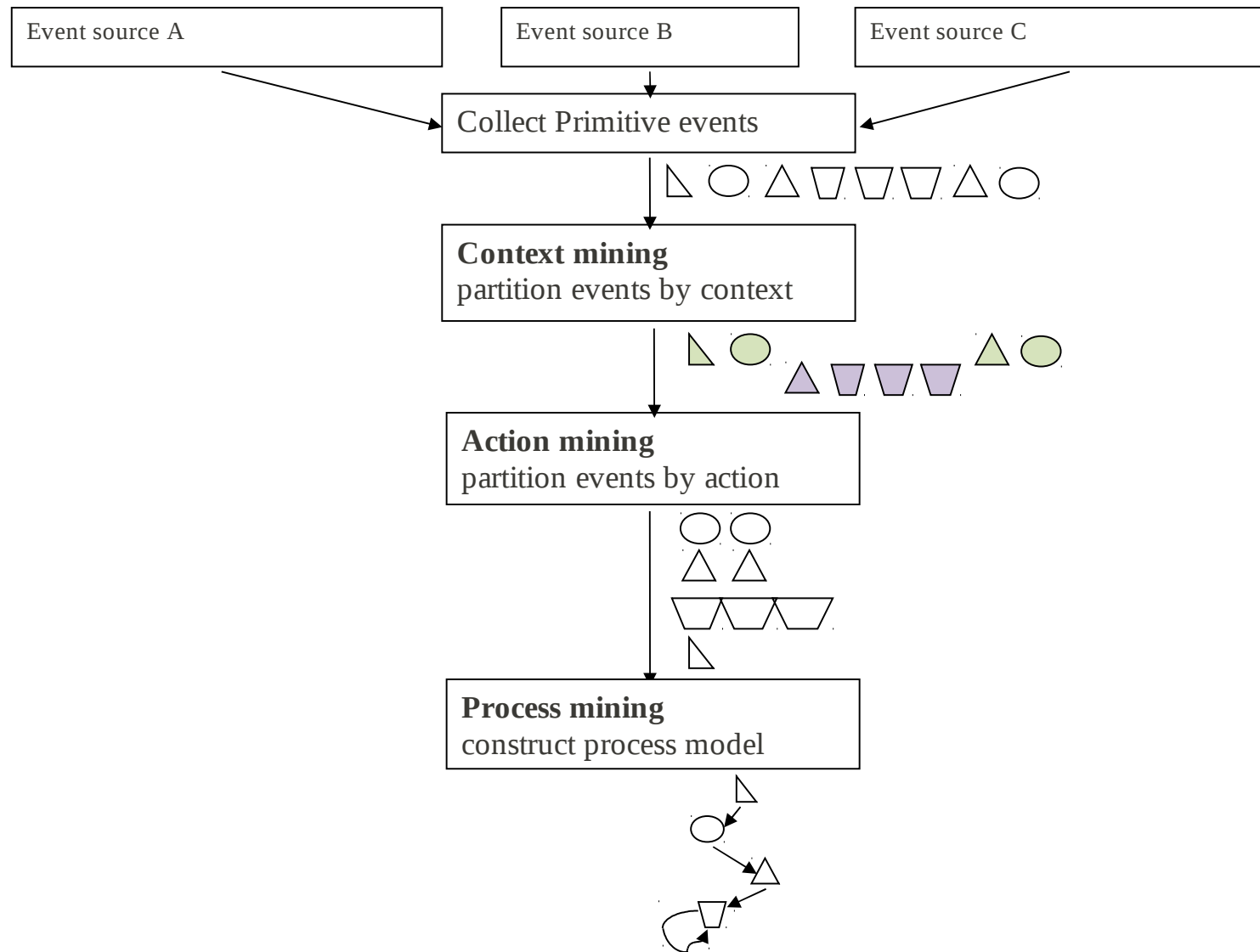
- ▶ To optimize the process, we need to be able to
 - ▶ Capture workflow logs
 - ▶ Construct a process model
 - ▶ Calculate optimization metrics
 - ▶ Refactor the actual process
 - ▶ Re-capture workflow log, measure improvement

Process model

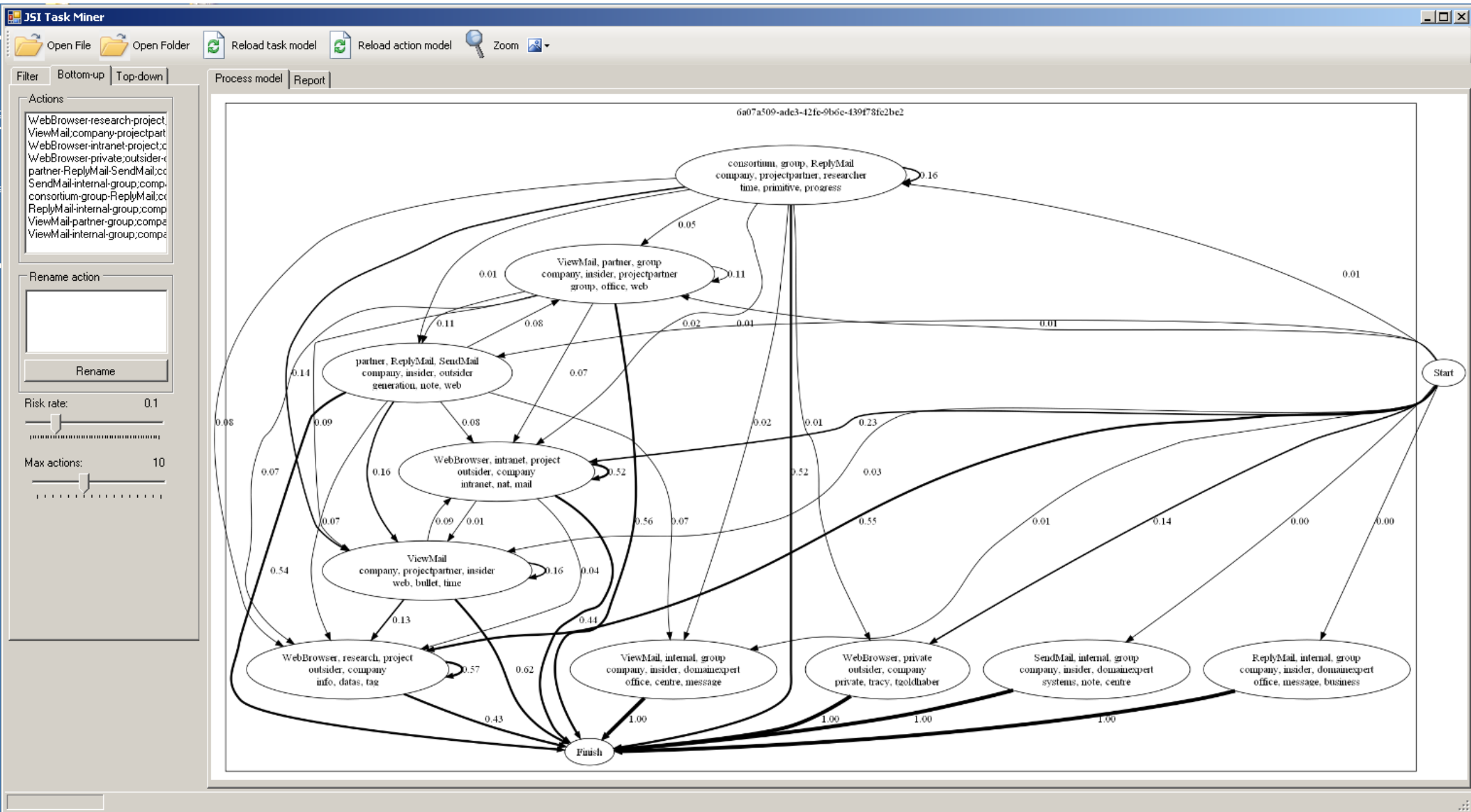




How it fits together



Process Visualization



Process Metric Reporting

JSI Task Miner

Open File Open Folder Reload task model Reload action model Zoom

Filter Bottom-up Top-down Process model Report

Actions

- WebBrowser-research-project;
- ViewMail;company-projectpart
- WebBrowser-intranet-project;c
- WebBrowser-private;outsider-c
- partner-ReplyMail-SendMail;cc
- SendMail-internal-group;comp
- consortium-group-ReplyMail;cc
- ReplyMail-internal-group;comp
- ViewMail-partner-group;compe
- ViewMail-internal-group;compe

Rename action

Rename

Risk rate: 0.1

Max actions: 10

Action	Total duration	Average duration	Time percentage	No. of events
ViewMail-partner-group;company-insider-projectpartner;group-office-web;	05:06:46	00:02:24.9290000	0.0383618	127
partner-ReplyMail-SendMail;company-insider-outsider;generation-note-web;	03:13:33	00:02:27	0.0242038	79
WebBrowser-intranet-project;outsider-company;intranet-nat-mail;	1:10:50:54	00:00:34.5410000	0.261471	3632
SendMail-internal-group;company-insider-domainexpert;systems-note-centre;	02:11:03	00:02:34.1760000	0.0163881	51
ViewMail-internal-group;company-insider-domainexpert;office-centre-message;	04:26:48	00:02:02.1980000	0.0333639	131
ReplyMail-internal-group;company-insider-domainexpert;office-message-business;	01:45:57	00:01:47.7450000	0.0132493	59
ViewMail;company-projectpartner-insider;web-bullet-time;	15:17:22	00:02:15.5710000	0.114719	406
WebBrowser-research-project;outsider-company;info-datas-tag;	2:05:34:28	00:00:21.0370000	0.401976	9168
consortium-group-ReplyMail;company-projectpartner-researcher;time-primitive-progress;	05:22:30	00:02:46.8100000	0.0403293	116
WebBrowser-private;outsider-company;private-tracy-tgoldhaber;	07:27:19	00:00:12.0670000	0.0559379	2224
Start	00:00:00	00:00:00	0	0
Finish	00:00:00	00:00:00	0	0

Conclusions and future work

- ▶ Process models of discovered actions have predictive power, not adversely affected by pruning, which allows for compromise on interpretability over predictive power
- ▶ Incorporating user input into context and action modeling
- ▶ End-user applications for making life easier for knowledge workers by providing context-dependent functionality and proactively suggesting relevant resources
- ▶ Manager-facing applications for observing process dynamics within their teams to provide infrastructure for process optimization