Chapter 13
Analyzing "Lasagna Processes"

Lasagna processes are relatively structured and the cases flowing through such processes are handled in a controlled manner. Therefore, it is possible to apply all of the process mining techniques presented in the preceding chapters. This chapter characterizes Lasagna processes and discusses typical use cases for process mining. Moreover, the different stages of a process mining project for improving a Lasagna process are described. The resulting life-cycle model guides users of process mining tools like ProM. Moreover, different application scenarios are discussed.

13.1 Characterization of "Lasagna Processes"

Unlike Spaghetti processes, Lasagna processes have a clear structure and most cases are handled in a prearranged manner. There are relatively few exceptions and stakeholders have a reasonable understanding of the flow of work. It is impossible to define a formal requirement characterizing Lasagna processes. As a rule of thumb we use the following informal criterion: *a process is a Lasagna process if with limited efforts it is possible to create an agreed-upon process model that has a fitness of at least 0.8*, i.e., more than 80% of the events happen as planned and stakeholders confirm the validity of the model. This implies (assuming that a suitable event log can be extracted) that all of the process mining techniques presented in this book can be applied.

The spectrum ranging from Lasagna processes to Spaghetti processes is a *continuum*. Sometimes the terms "structured", "semi-structured", and "unstructured" are used to refer to the same continuum. In a *structured process* (i.e., Lasagna process) all activities are repeatable and have a well defined input and output. In highly structured processes most activities can, in principle, be automated. In *semistructured processes* the information requirements of activities are known and it is possible to sketch the procedures followed. However, some activities require human judgment and people can deviate depending on taste or the characteristics of the case being handled. In *unstructured processes* (i.e., Spaghetti process) it is difficult to define

W. van der Aalst, *Process Mining*, DOI 10.1007/978-3-662-49851-4_13

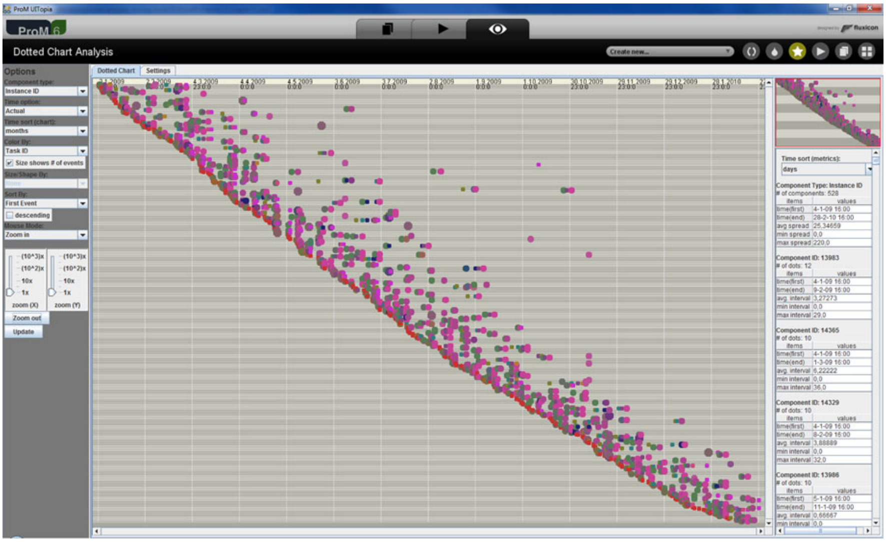


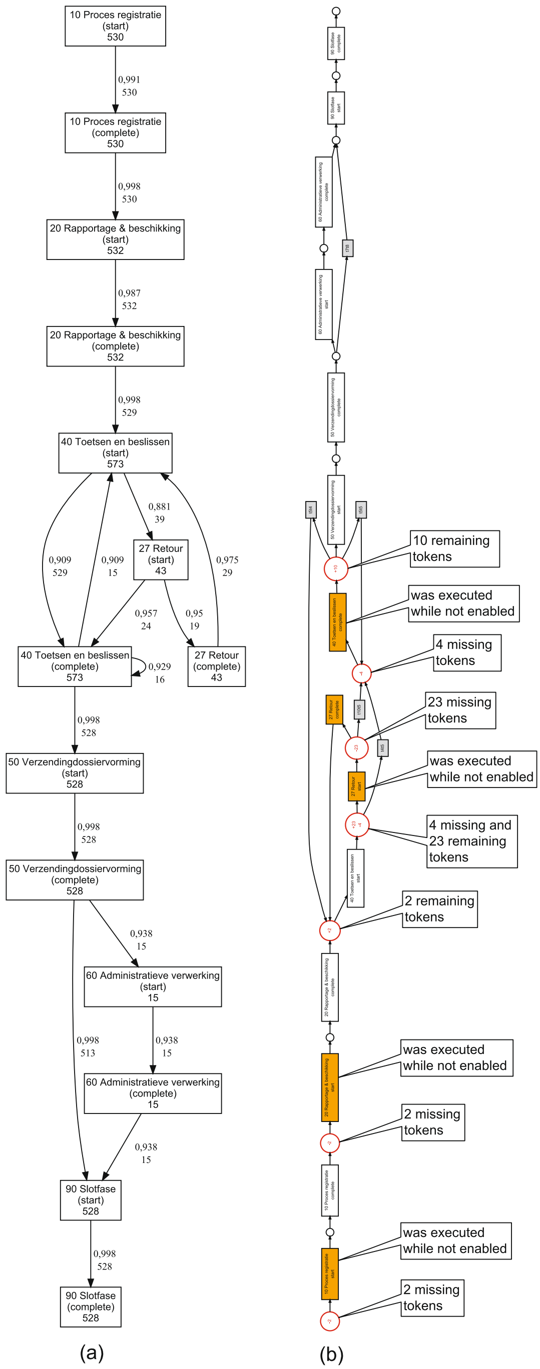
Fig. 13.1 Screenshot of ProM 6 showing a dotted chart for a WMO process of a Dutch municipality. Each line corresponds to one of the 528 requests that were handled in the period from 4-1-2009 until 28-2-2010. In total there are 5498 events represented as dots. The mean time needed to handled a case is approximately 25 days

pre- and post-conditions for activities. These processes are driven by experience, intuition, trail-and-error, rules-of-thumb, and vague qualitative information.

Let us consider an example of a Lasagna process. Figure 13.1 shows a dotted chart for one of the so-called WMO processes of a Dutch municipality. WMO (Wet Maatschappelijke Ondersteuning) refers to the social support act that came into force in The Netherlands on January 1st, 2007. The aim of this act is to assist people with disabilities and impairments. Under the act, local authorities are required to give support to those who need it, e.g., household help, providing wheelchairs and scootmobiles, and adaptations to homes. There are different processes for the different kinds of help. The dotted chart in Fig. 13.1 is based on the process for handling requests for household help. In a period of about one year, 528 requests for household WMO support were received. These 528 requests generated 5498 events each represented as a colored dot in Fig. 13.1. The color of the dot refers to the activity executed for the request, e.g., a red dot refers to activity “10 Process registratie” (register request) and a blue dot refers to activity “40 toetsen en beslissen” (evaluate and decide). The diagonal line of initial events shows that there is a steady flow of new requests. The dots also show that the time to completely handle requests is typically short (about one month).

Although no process model is shown in Fig. 13.1, the dotted chart already suggests that the process is a Lasagna process (regular arrival pattern, most cases are handled within one month, and clearly noticeable recurring patterns). Figure 13.2 demonstrates that this is indeed the case. The process model discovered by the heuristic miner shows that the process is highly structured and rather sequential.

Fig. 13.2 The C-net discovered using the heuristic miner (**a**) and the corresponding Petri net with missing and remaining tokens after replay (**b**). The numbers generated by the heuristic miner show the flow of tokens. The C-net was translated into an equivalent Petri net with silent transitions. The fitness was analyzed using ProM's conformance checker (cf. Sect. 8.2). The fitness of the discovered process is 0.99521667. Of the 528 cases, 496 cases fit perfectly whereas for 32 cases there are missing or remaining tokens. The missing and remaining tokens show where the model and log deviate. For example, for two cases the activity "40 toetsen en beslissen" (evaluate and decide) was not started although it should have. Activity "20 Rapportage & beschikking" (report and intermediate decision) was started twice while this was not possible according to the model



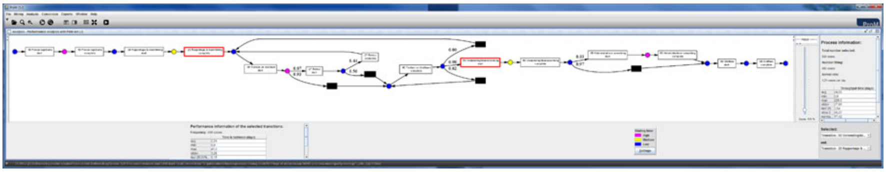


Fig. 13.3 Screenshot of ProM 5.2 while analyzing the bottlenecks in the process. The mean flow time of fitting cases is 24.66 days. Most time is spent on the activities "10 Process registratie", "40 Toetsen en beslissen", and "60 Administratieve verwerking". The average time in-between the completion of activity "10 Rapportage & beschikking" and "50 Verzending/dossiervorming" is 2.24 days

The figure does not show the logic of splits and joins, e.g., one cannot see the difference between AND/OR/XOR-splits/joins.[1] ProM's heuristic miner does not allow for the visualization of bindings used in Sect. 7.2. However, the logic of splits and joins is also discovered and can be shown if desired. When converting a C-net into a Petri net, EPC model, of BPMN model this information is taken into account. The discovered C-net in Fig. 13.2(a) is annotated with frequencies. The frequency of a node indicates how often the corresponding activity appeared in the event log. For instance, activity "20 Rapportage & beschikking" (report and intermediate decision) occurred 532 times. Arcs have a frequency indicating how often a token was passed along the arc when replaying the log. Figure 13.2(b) shows a WF-net obtained by using the corresponding conversion plug-in in ProM. The conformance checker of ProM shows that the fitness of model and log is 0.99521667. This shows that there are hardly any missing or remaining tokens when replaying all 528 cases. Figure 13.2(b) also shows some of the detailed diagnostics. The discovered process model and the high fitness value show that the WMO process is definitely a Lasagna process. This implies that, in principle, *all process mining techniques described in this book are applicable to this process* (assuming sufficient event data). Figure 13.3 shows one of many process mining techniques that can be applied. As explained in Sect. 9.4, delays can be analyzed by replaying the event log while taking timestamps into account. Figure 13.3 illustrates that it is possible to discover bottlenecks for a Lasagna process like the WMO process. Note that the plug-in used in Fig. 13.3 exploits the coupling between the event log and the discovered model (cf. Fig. 13.2).

In Sect. 13.4, we provide more examples of Lasagna processes. However, first we discuss typical use cases for process mining and present a life-cycle model for process mining projects.

[1] In the remainder, we will never show the set of input and output bindings for C-nets discovered by the heuristic miner. The heuristic miner can visualize the logic of splits and joins, but this typically impairs the readability of the diagram.

13.2 Use Cases

The goal of process mining is to *improve* operational processes. In order to judge whether process mining efforts are successful, we need to define *Key Performance Indicators* (KPIs). In Sect. 3.3.2, we identified three classes of KPIs: KPIs related to *time* (e.g., lead time, service time, waiting time, and synchronization time), KPIs related to *costs*, and KPIs related to *quality*. Note that quality may refer to compliance, customer satisfaction, number of defects, etc. To evaluate suggested improvements, the effectiveness and efficiency of the *as-is* and *to-be* processes need to be quantified in terms of KPIs.

For Lasagna processes, process mining can result in one or more of the following *improvement actions*:

- *Redesign.* Insights obtained using process mining can trigger changes to the process, e.g., sequential activities no longer need to be executed in a fixed order, checks may be skipped for easy cases, decisions can be delegated if more than 50 cases are queueing, etc. Fraud detected using process mining may result in additional compliance regulations, e.g., introducing the 4-eyes principle for critical activities.
- *Adjust.* Similarly, process mining can result in (temporary) adjustments. For example, insights obtained using process mining can be used to temporarily allocate more resources to the process and to lower the threshold for delegation.
- *Intervene.* Process mining may also reveal problems related to particular cases or resources. This may trigger interventions such as aborting a case that has been queuing for more than 3 months or disciplinary measures for a worker that repeatedly violated compliance regulations.
- *Support.* Process mining can be used for operational support, e.g., based on historic information a process mining tool can predict the remaining flow time or recommend the action with the lowest expected costs.

Figure 2.4 in Chap. 2 illustrates the difference between a *redesign* (a permanent change requiring alterations to software or model) and an *adjustment* (a temporary change realized without modifying the underlying software or model).

As shown in Fig. 13.4, *use cases for process mining refer to a combination of KPIs and improvement actions.* Given a Lasagna process, some typical use cases for process mining are:

- Identification of bottlenecks to trigger a process redesign that reduces the overall flow time with 30%.
- Identification of compliance problems using conformance checking. Some of the compliance problems result in ad-hoc interventions whereas others lead to adjustments of the parameters used for work distribution.
- Harmonization of two processes after a merger based on a comparison of the actual processes. The goal of such a harmonization is to reduce costs.
- Predicting the remaining flow time of delayed cases to improve customer service.
- Providing recommendations for resource allocation aiming at a more balanced utilization of workers.

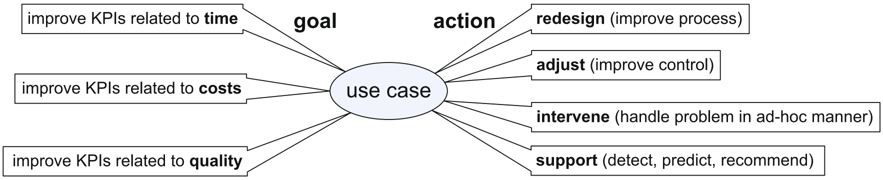


Fig. 13.4 Use cases for process mining combine goals (expressed in KPIs) and improvement actions, e.g., process mining can be used to shorten the flow time by providing insights that lead to a process redesign

- Identification of exceptional cases that generate too much additional work. By learning the profile of such cases, they can be handled separately to reduce the overall flow time.
- Visualization of the 10 most complicated or time consuming cases to identify potential risks.

These use cases illustrate the potential of process mining. It is easy to imagine the application of these use cases to the WMO process described earlier. For instance, results such as shown in Fig. 13.3 can be used to discover bottlenecks and to generate ideas for flow time reduction. The results of conformance analysis as depicted in Fig. 13.2(b) can be used to identify compliance problems, e.g., for the 32 cases having missing or remaining tokens one could analyze the social network of the people involved.

13.3 Approach

In Chap. 10, we described ten process mining related activities using the framework shown in Fig. 13.5. These ten activities are grouped into three categories: cartography (activities *discover*, *enhance*, and *diagnose*), auditing (activities *detect*, *check*, *compare*, and *promote*), and navigation (activities *explore*, *predict*, and *recommend*). Although the framework helps to understand the relations between the various process mining activities, it does not guide the user in conducting a process mining project. Therefore, this section introduces the *L^* life-cycle model for mining Lasagna processes*.

Several reference models describing the life-cycle of a typical data mining/BI project have been proposed by academics and consortia of vendors and users. For example, the *CRISP-DM* (CRoss-Industry Standard Process for Data Mining) methodology identifies a life-cycle consisting of six phases: (a) business understanding, (b) data understanding, (c) data preparation, (d) modeling, (e) evaluation, and (f) deployment [29]. CRISP-DM was developed in the late nineties by a consortium driven by SPSS. Around the same period SAS proposed the *SEMMA* methodology consisting of five phases: (a) sample, (b) explore, (c) modify, (d) model, and (e) assess. Both methodologies are very high-level and provide little support. Moreover, existing methodologies are not tailored towards process mining projects. Therefore, we propose the L^* life-cycle model shown in Fig. 13.6. This five-stage model

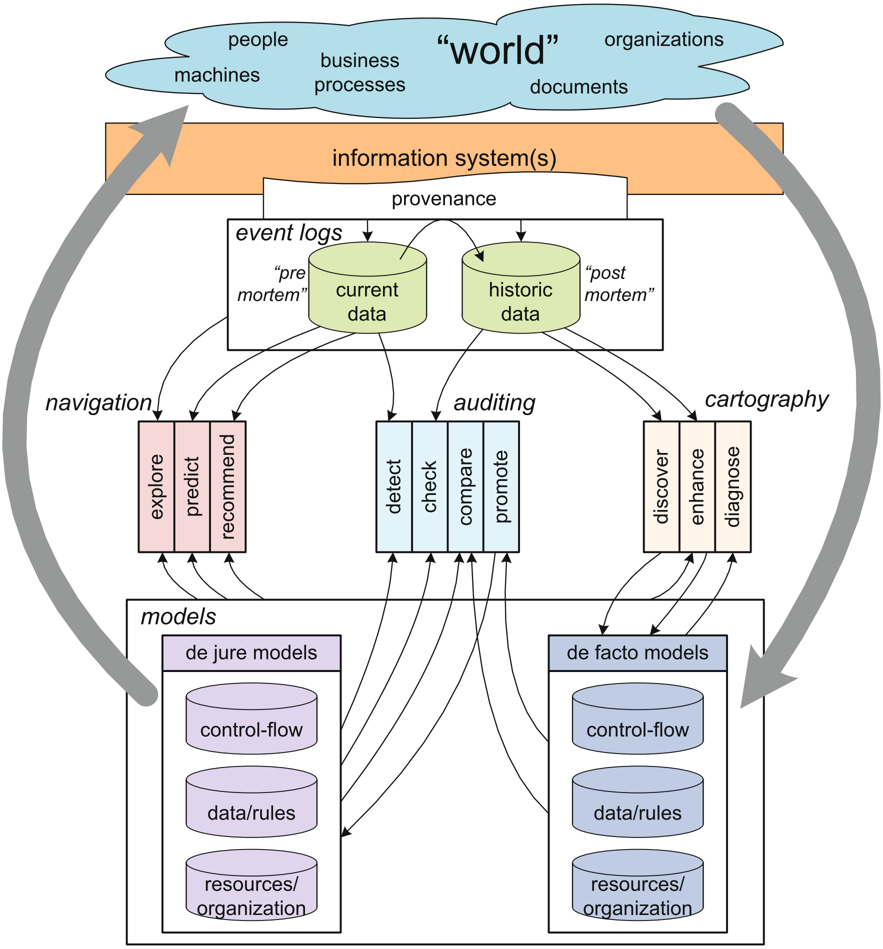


Fig. 13.5 The process mining framework introduced in Chap. 10. The framework identifies ten process mining activities (discover, check, enhance, etc.)

describes the life-cycle of a typical process mining project aiming to improve a Lasagna process.

In the remainder, we discuss each of the five stages. As shown in Fig. 13.6, the L^* life-cycle model refers to the ten process mining related activities (explore, discover, check, etc.) and the four improvement actions (redesign, adjust, intervene, and support) mentioned earlier.

13.3.1 Stage 0: Plan and Justify

Any process mining project starts with a planning and a justification of the planned activities. Before spending efforts on process mining activities, one should antic-

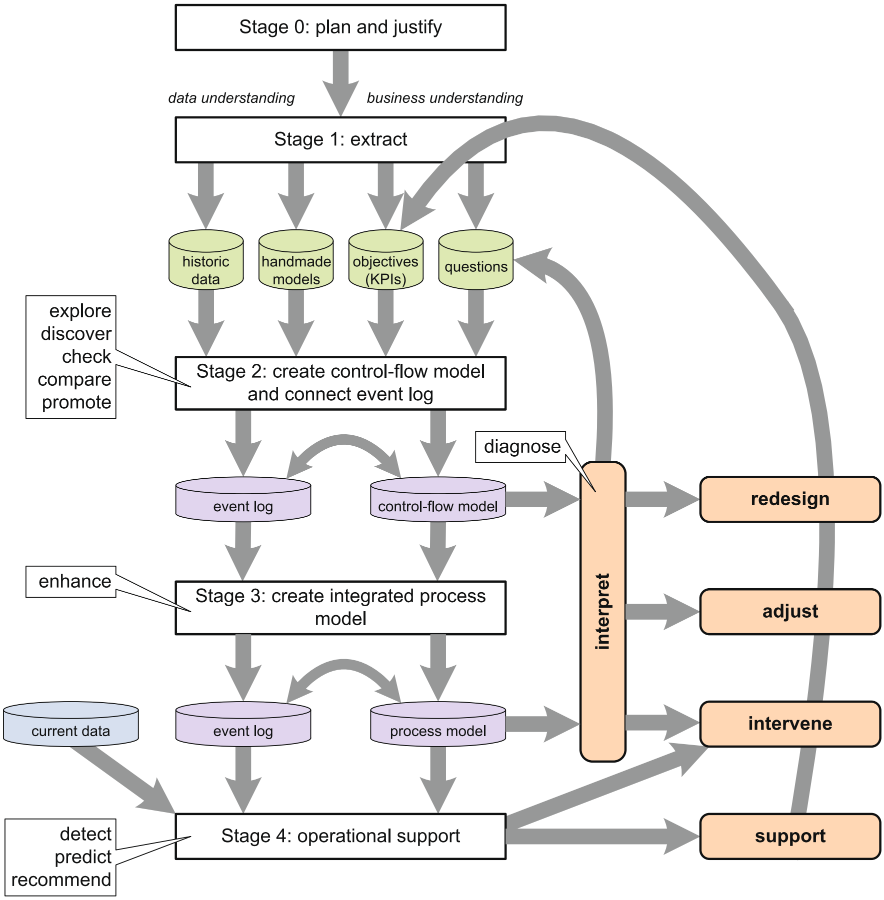


Fig. 13.6 The L^* life-cycle model describing a process mining project consisting of five stages: *plan and justify* (Stage 0), *extract* (Stage 1), *create control-flow model and connect event log* (Stage 2), *create integrated process model* (Stage 3), and *operational support* (Stage 4)

ipate benefits that may result from the project. There are basically three types of process mining projects:

- A *data-driven* (also referred to as “curiosity driven”) process mining project is powered by the availability of event data. There is no concrete question or goal, however, some of the stakeholders expect that valuable insights will emerge by analyzing event data. Such a project has an explorative character.
- A *question-driven* process mining project aims to answer specific questions, e.g., “Why do cases handled by team X take longer than cases handled by team Y?” or “Why are there more deviations in weekends?”.
- A *goal-driven* process mining project aspires to improve a process with respect to particular KPIs, e.g., cost reduction or improved response times.

For an organization without much process mining experience it is best to start with a question-driven project. Concrete questions help to scope the project and guide data extraction efforts.

Like any project, a process mining project needs to be planned carefully. For instance, activities need to be scheduled before starting the project, resources need to be allocated, milestones need to be defined, and progress needs to be monitored continuously.

13.3.2 Stage 1: Extract

After initiating the project, event data, models, objectives, and questions need to be extracted from systems, domain experts, and management.

In Chap. 5, we elaborated on data extraction. For example, Fig. 5.1 describes the process of getting from raw data to suitable event logs. Recall that event logs have two main requirements: (a) events need to be ordered in time and (b) events need to be correlated (i.e., each event needs to refer to a particular case).

As Fig. 13.6 shows, it is possible that there are already handmade (process) models. These models may be of low quality and have little to do with reality. Nevertheless, it is good to collect all models present and exploit existing knowledge as much as possible. For example, existing models can help in scoping the process and judging the completeness of event logs.

In a goal-driven process mining project, the objectives are also formulated in Stage 1 of the L^* life-cycle. These objectives are expressed in terms of KPIs. In a question-driven process mining project, questions need to be generated in Stage 1. Both questions and objectives are gathered through interviews with stakeholders (e.g., domain experts, end users, customers, and management).

13.3.3 Stage 2: Create Control-Flow Model and Connect Event Log

Control-flow forms the backbone of any process model. Therefore, Stage 2 of the L^* life-cycle aims to determine the de facto control-flow model of the process that is analyzed. The process model may be discovered using the process discovery techniques presented in Part III of this book (activity *discover* in Fig. 13.6). However, if there is a good process model present, it may be verified using conformance checking (activity *check*) or judged against the discovered model (activity *compare*). It is even possible to merge the handmade model and the discovered model (activity *promote*). After completing Stage 2 there is a control-flow model tightly connected to the event log, i.e., events in the event log refer to activities in the model. As discussed in Sect. 8.5.3, this connection is crucial for subsequent steps. If the fitness of the model and log is low (say below 0.8), then it is difficult to move to Stage 3. However, by definition, this should not be a problem for a Lasagna process.

The output of Stage 2 may be used to answer questions, take actions, or to move to Stage 3. As Fig. 13.6 shows, the output (control-flow model connected to an event log) needs to be interpreted before it can be used to answer questions or trigger a redesign, an adjustment, or an intervention.

13.3.4 Stage 3: Create Integrated Process Model

In Stage 3, the model is enhanced by adding additional perspectives to the control-flow model (e.g., the organizational perspective, the case perspective, and the time perspective). Chapter 9 shows how these perspectives can be discovered and integrated, e.g., Fig. 9.16 describes the process of merging the different perspectives. The result is an integrated process model that can be used for various purposes. The model can be inspected directly to better understand the as-is process or to identify bottlenecks. Moreover, a complete process model can also be simulated as discussed in Sect. 9.6.

The output of Stage 3 can also be used to answer selected questions and take appropriate actions (redesign, adjust, or intervene). Moreover, the integrated process model is also input for Stage 4.

13.3.5 Stage 4: Operational Support

Stage 4 of the L^* life-cycle is concerned with the three operational support activities described in Chap. 10: *detect*, *predict*, and *recommend*. For instance, using short-term simulation (Sect. 9.6) or annotated transition systems (Sect. 10.4) it is possible to predict the remaining flow time for running cases. As shown in Fig. 13.6, Stage 4 requires current data ("pre mortem" data on running cases) as input. Moreover, the output does not need to be interpreted by the process mining analyst and can be directly offered to end users. For example, a deviation may result in an automatically generated e-mail sent to the responsible manager. Recommendations and predictions are presented to the persons working on the corresponding cases.

Note that operational support is the *most ambitious* form of process mining. This is only possible for Lasagna processes. Moreover, there needs to be an advanced IT infrastructure that provides high-quality event logs and allows for the embedding of an operational support system as described in Chap. 10.

The PM^2 process mining methodology presented in [175] can be viewed as a refinement of the L^* life-cycle. Using a case study conducted within IBM, the PM^2 methodology is explained. Moreover, selected ProM plug-ins are related to the different phases in [175].

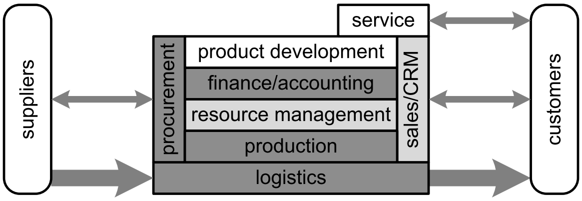


Fig. 13.7 Overview of the different functional areas in a typical organization. Lasagna processes are typically encountered in production, finance/accounting, procurement, logistics, resource management, and sales/CRM. Spaghetti processes are typically encountered in product development, service, resource management, and sales/CRM

13.4 Applications

In the last decade, we have applied process mining in over 150 organizations. Examples are municipalities (e.g., Alkmaar, Heusden, and Harderwijk), government agencies (e.g., Rijkswaterstaat, Centraal Justitieel Incasso Bureau, and Justice department), insurance related agencies (e.g., UWV), banks (e.g., ING Bank), hospitals (e.g., AMC hospital and Catharina hospital), multinationals (e.g., DSM and Deloitte), high-tech system manufacturers and their customers (e.g., Philips Healthcare, ASML, Ricoh, and Thales), and media companies (e.g., Winkwaves). This illustrates the broad spectrum of situations in which process mining can be applied. In remainder of this section, we identify process mining opportunities in different functional areas and in different sectors and industries. Moreover, we briefly discuss two case studies involving Lasagna processes.

13.4.1 Process Mining Opportunities per Functional Area

Figure 13.7 shows the main *functional areas* that can be found in most organizations:

- *Product development* is concerned with all the preparations and engineering work needed to start producing a particular product. Products do not need to be physical objects (e.g., a car or copier); the product may also be a piece of information or a service (e.g., a new kind of insurance). Product development processes are typically Spaghetti-like because they have a lower frequency and depend on problem solving, expertise, and creativity rather than repetition, routine, and efficiency.
- *Production* is the functional area where the products are actually produced. Processes may range from classical manufacturing (assembling a car) to information creation (opening a back account). Most production processes are Lasagna processes because they need to be reproducible and efficient.
- *Procurement* entails all activities to get the materials needed for production. Note that the input for the production process may also be information from other parties. The input materials need to be purchased, stocks need to be monitored,

deliveries need to be checked, etc. Processes in this functional area are typically Lasagna processes.

- The functional area *Sales/CRM* is concerned with all activities related to "lead-to-order" and "order-to-cash". Besides the actual sales function, most organizations need to market their products and manage long-term relationships with their customers (CRM). Both Lasagna processes and Spaghetti processes can be found in this functional area. The handling of sales activities can be very structured whereas marketing-related activities may be rather unstructured.
- *Logistics* is concerned with the movements of products and materials, e.g., shipping the product to the customer and managing the storage space. Most processes in logistics are Lasagna processes.
- The functional area *Finance/accounting* deals with all financial aspects of an organization, e.g., billing customers, checking invoices, financial reporting, and auditing. Processes in this functional area are also typically Lasagna processes.
- *Resource management* is the functional area that makes sure there are sufficient resources to perform all other functions. HRM (Human Resource Management) is concerned with human resources and similar functions exist for machines, buildings, etc. Both Lasagna processes and Spaghetti processes can be found in this functional area, e.g., the handling of job applications may be very structured whereas the handling of a problematic employee may be rather ad-hoc.
- The functional area *Service* deals with all activities after the product has been shipped and paid for, e.g., activities related to product support, maintenance, repairing defective products, and help-desk operations. Service related processes are typically Spaghetti-like. Customers will use products in many different ways and repair processes are rather unpredictable for most products, e.g., no faults are found in the product returned by the customer or the wrong component is replaced and the product still malfunctions intermittently.

The characterization of the different functional areas in terms of Lasagna processes and Spaghetti processes is only intended as an indication. Both types of processes can be found in all of the functional areas. However, as shown in Fig. 13.7, it is possible to pinpoint typical functional areas for both types. For example, in most organizations product development processes are rather unstructured compared to production processes. This implies that most of the techniques presented in this book can be applied to production processes. However, for product development processes it is unlikely that all stages of the L^* life-cycle model (Fig. 13.6) can be executed. (Stages 3 and 4 are typically not possible for Spaghetti-like processes.)

13.4.2 Process Mining Opportunities per Sector

After contemplating on the presence of Lasagna and Spaghetti processes in the functional areas in one organization (Fig. 13.7), we now look at different sectors and industries.

The *primary sector* of the economy is concerned with transforming natural resources into primary products (e.g., agriculture, agribusiness, fishing, forestry and all mining and quarrying industries). Information technology tends to play a minor role in these industries. Hence, the application potential of process mining is limited. Of course there are exceptions. Consider for instance the tracking and tracing of food. In some countries meat and dairy products need to be tracked from source to sink. For example, meat products in supermarkets need to be linked to particular animals and farms. This requires the recording of events starting in the primary sector.

The *secondary sector* of the economy refers to the manufacturing of tangible products and includes the automotive industry, chemical industry, aerospace manufacturing, consumer electronics, etc. Organizations in the secondary sector typically have an organizational structure covering all functional areas depicted in Fig. 13.7. Hence, both Lasagna processes and Spaghetti processes can be encountered. An interesting observation across the different industries is that most manufacturers have become interested in monitoring their products after they have been sold. For example, Philips Healthcare is monitoring their medical equipment while being deployed in the field, e.g., their X-ray machines are connected to the Internet and the resulting logs are analyzed using ProM. The event logs of these X-ray machines provide vital information for marketing (What kind of features do customer use?), maintenance (When to service the machine?), development (Why do machines fail?), and testing (How to test machines under realistic circumstances?). In the future, more and more (consumer) products will be monitored remotely thus providing valuable information for the manufacturer.

The *tertiary sector* of the economy consists of all organizations that produce "intangible goods" such as services, regulations, and information. The term "services" should be interpreted in the broadest sense including transportation, insurance, wholesaling, retailing, entertainment, etc. Note that goods may be transformed in the process of providing the service (cf. preparing food in a restaurant). However, the focus is on serving the customer rather than transforming physical goods. In many industries in the tertiary sector, information plays a dominant role and many events are being recorded. This is the sector where the digital universe and the physical universe are aligned most. For example, an electronic bookstore can only sell a book if the information system indicates that the book is present. The bookstore would not be able to sell a particular book if the information system would indicate that it is out-of-stock; even if the book would be physically present in the warehouse.

Process mining can be used to improve a variety of Lasagna and Spaghetti processes encountered in the tertiary sector. Below we sketch some of the most interesting industries.

- The *healthcare* industry includes hospitals and other care organizations. Most events are being recorded (blood tests, MRI scans, appointments, etc.) and correlation is easy because each event refers to a particular patient. The closer processes get to the medical profession, the less structured they become. For instance, most diagnosis and treatment processes tend to be rather Spaghetti-like (see Fig. 14.1). Medical guidelines typically have little to do with the actual

processes. On the one hand, this suggests that these processes can be improved by structuring them. On the other hand, the variability of medical processes is caused by the different characteristics of patients, their problems, and unanticipated complications. Patients are saved by doctors deviating from standard procedures. However, some deviations also cost lives. Clearly, hospitals need to get a better understanding of care processes to be able to improve them. Process mining can help as event data is readily available [95].
- *Governments* range from small municipalities to large organizations operating at the national level, e.g., institutions managing processes related to unemployment, customs, taxes, and traffic offences. Both local and national government agencies can be seen as "administrative factories" as they execute regulations and the "products" are mainly informational or financial. Processes in larger government agencies are characterized by a high degree of automation. Consider, for example, tax departments that need to deal with millions of tax declarations. Processes in smaller government agencies (e.g., small municipalities) are typically not automated and managed by office workers rather than BPM systems. However, due to the legal requirements, all main events are recorded in a systematic manner. Consider, for example, the WMO process shown in Fig. 13.2; any municipality in The Netherlands is obliged to record the formal steps in such processes. Typical use cases for process mining in governments (local or non-local) are flow time reduction (e.g., shorten the time to get a building permit), improved efficiency, and compliance. Given the role of governments in society, compliance is of the utmost importance.
- *Banking* and *insurance* are two industries where BPM technology has been most effective. Processes are often automated and all events are recorded in a systematic and secure manner. Examples are the processing of loans, claims management, handling insurance applications, credit card payments, and mortgage payments. Most processes in banking and insurance are Lasagna processes, i.e., highly structured. Hence, all of the techniques presented in this book can be applied. Process discovery is less relevant for these organizations as most processes are known and documented. Typical uses cases in these industries involve conformance checking, performance analysis, and operational support.
- Organizations involved in *education* (e.g., high-schools and universities) are recording more and more information related to the study behavior of individuals. For instance, at TU/e we are applying process mining to analyze study behavior using a database containing detailed information about exam results of all students that ever studied computer science. Moreover, this database also contains information about high-school exam grades, etc. Some of these educational processes are structured, others are very unstructured. For example, it is very difficult to predict the remaining study time of students at a university because the curriculum often changes and students tend to have very different study patterns. Nevertheless, valuable insights can be obtained. By visualizing that few students follow the courses in the order intended, one can show that the design of a curriculum should not only focus on the "ideal student" (that passes all courses the first time), but also anticipate problems encountered by other students.

- The products manufactured by organizations in the secondary sector are distributed through various *retail* organizations. Here it is interesting to see that more and more information about products and customers is being recorded. Customers are tracked using loyalty cards or through online profiles. Products are tagged and the shop has real-time information about the number of items still available. A product that has an RFID tag has a unique identifier, i.e., two identical products can still be distinguished. This allows for the correlation of events and thus facilitates process mining.
- The *transportation* industry is also recording more and more information about the movement of people and products. Through tracking and tracing functionality the whereabouts of a particular parcel can be monitored by both sender and receiver. Although controversial, smartcards providing access to buildings and transportation systems can be used to monitor the movement of people. For example, the Dutch "ov-chipkaart" can be used to travel by train, subway, and bus. The traveler pays based on the distance between the entry point and exit point. The recorded information can be used to analyze traveling behavior. The booking of a flight via the Internet also generates lots of event data. In fact, the booking process involves only electronic activities. Note that the traveler interacts with one organization that contacts all kinds of other organizations in the background (airlines, insurance companies, car rental agencies, etc.). All of these events are being recorded, thus enabling process mining. The whole spectrum ranging from Lasagna processes to Spaghetti processes can be found in this industry.
- New technologies such as *cloud computing* and *Software-as-a-Service* (SaaS) have created a new industry that offers computing as a utility (like water and electricity). Google Apps. Salesforce.com, and Amazon EC2/S3 are examples of companies providing such utilities. The idea is not new: already in 1961 John McCarthy stated "If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. The computer utility could become the basis of a new and important industry." A well-known example of a SaaS provider that is using a cloud infrastructure is SalesForce.com. This company allows organizations to outsource the IT support of standard activities related to sales and CRM without worrying about scalability and maintenance. Users pay for using the software rather than owning it. Another example is the conference management system EasyChair that is currently probably the most commonly used system to host conferences and to manage the reviewing of scientific papers. To organize a conference, there is no need to install any software as everything is hosted and managed centrally. Organizations such as SalesForce.com and EasyChair have access to valuable event data. These data can be used to improve their software and to give advice to individual organizations. One of the challenges SaaS providers are facing is the need to deal with variability across organizations. Process mining can help analyzing differences between organizations using *cross-organizational process mining*, i.e., using process mining to compare similar processes within the same or in different organizations.

- The *capital goods* industry is also transforming from the situation in which customers purchase expensive machines to the situation in which customers only pay for the actual use of the machine. Note that this can be seen as a variant of the SaaS paradigm. The manufacturer of the machine remains being the owner and customers pay depending on usage and uptime of the machine. Clearly, such pricing models require the remote monitoring of capital goods. For instance, service provider and consumer need to agree on the actual use (e.g., hours of use or number of production cycles). Moreover, there may be Service Level Agreements (SLAs) specifying a fine if the machine is down for an extended period. Event data can be used as a basis for billing and checking SLAs. Moreover, the manufacturer gets insights into the way that machines are used, when they malfunction, and when they require maintenance.

These examples show that there are opportunities for process mining in all three economic sectors.

13.4.3 Two Lasagna Processes

To conclude this chapter, we briefly discuss two case studies analyzing Lasagna processes.

13.4.3.1 RWS Process

The Dutch national public works department, called "Rijkswaterstaat" (RWS), has 12 provincial offices. We analyzed the handling of invoices in one of these offices [160]. The office employs about 1,000 civil servants and is primarily responsible for the construction and maintenance of the road and water infrastructure in its province. To perform its functions, the RWS office subcontracts various parties such as road construction companies, cleaning companies, and environmental bureaus. Also, it purchases services and products to support its construction, maintenance, and administrative activities. The reason to employ process mining within RWS was twofold. First of all, RWS was involved in our longitudinal study into the effectiveness of WFM systems [116]. In the context of this study, RWS was interested to see the effects of WFM technology on flow times, response times, service levels, utilization, etc. Second, RWS was interested in better meeting deadlines with respect to the payment of invoices. Payment should take place within 31 days from the moment the invoice is received. After this period, the creditor is entitled (according to Dutch law) to receive interest over the outstanding sum. RWS would like to pay at least 90% of its invoices within 31 days. However, analysis of the event logs of RWS showed that initially only 70% of payments were paid in time.

Starting point for the analysis described in [160] was an event log containing information about 14,279 cases (i.e., invoices) generating 147,579 events. Figure 13.8 shows a C-net generated by the heuristic miner. This model shows that the RWS

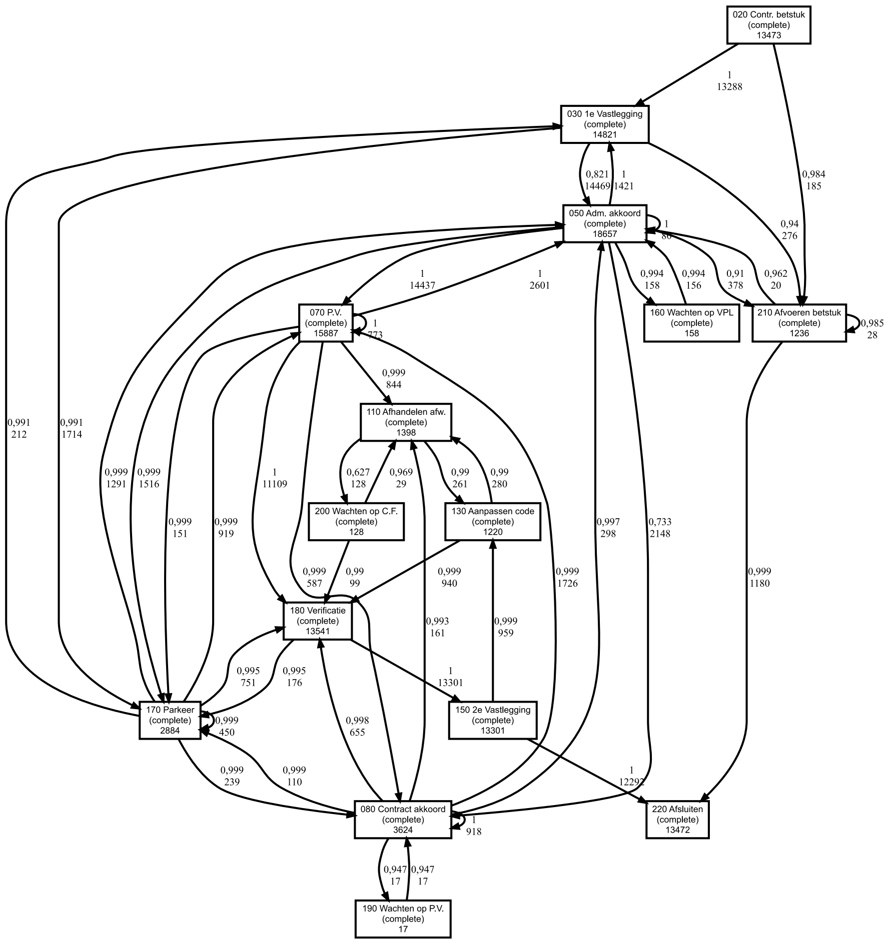


Fig. 13.8 Process model obtained using heuristic mining. The C-net describes the handling of invoices within one of the twelve provincial offices of RWS

process is fairly structured, but not as structured as the WMO process depicted in Fig. 13.2(a). After some efforts (filtering the log and tuning the parameters of the mining algorithm), it is possible to create a model with a fitness of more than 0.9. The log can be replayed on this model to highlight bottlenecks. Such analysis shows that several activities had to be re-done (as can be seen by the loops of length one or two in Fig. 13.8), i.e., work was sent "back-and-forth" between different activities and people thus causing delays.

The event log contains information about 271 resources, i.e., civil servants involved in the handling invoices. Figure 13.9 shows the social network based on the frequency of handovers (cf. Sect. 9.3.1). Figure 13.10 shows the same social network, but now only for the 13 resources that executed most activities. RWS could

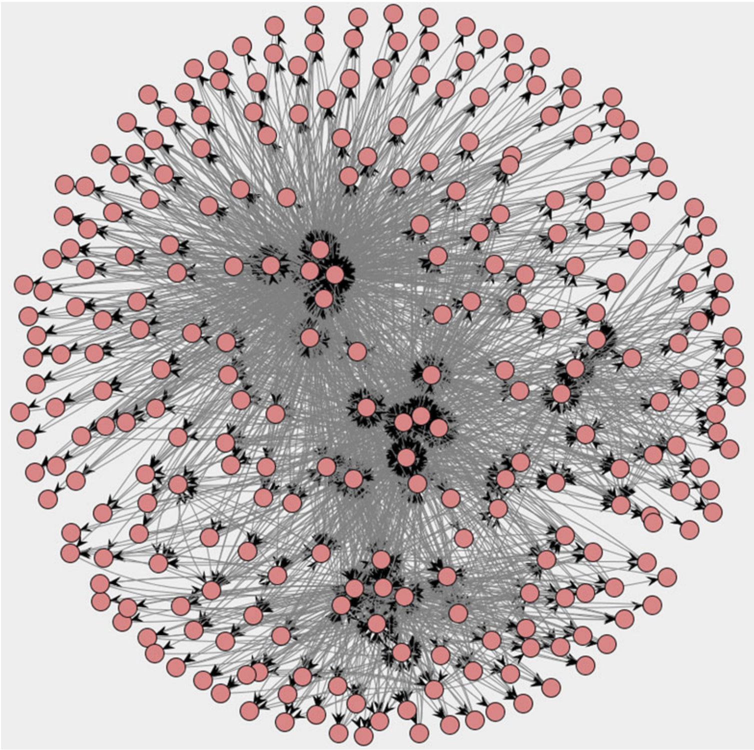


Fig. 13.9 Social network constructed based on handovers of work. Each of the 271 nodes corresponds to a civil servant. Two civil servants are connected if one executed an activity causally following an activity executed by the other civil servant

use these social networks to better understand how work is flowing through the organization. This analysis showed that some project leaders considered invoice approval to be of low priority, not realizing that because of their slow reaction time many invoices took more than 31 days. They were not aware of the impact of their actions and agreed to give the invoice approval a higher priority thus speeding up the process. See [160] for more information.

13.4.3.2 WOZ Process

In Sect. 13.1, we showed some analysis results for a WMO process of a municipality. To date, we have applied process mining in about a dozen municipalities. Moreover, we just started a new project (CoSeLoG) involving nine municipalities interested in cross-organizational process mining, i.e., analyzing differences between similar processes in different municipalities [35].

Processes in municipalities are typically Lasagna processes. To illustrate this we present another example. Figure 13.11 shows a so-called "WOZ process" discov-

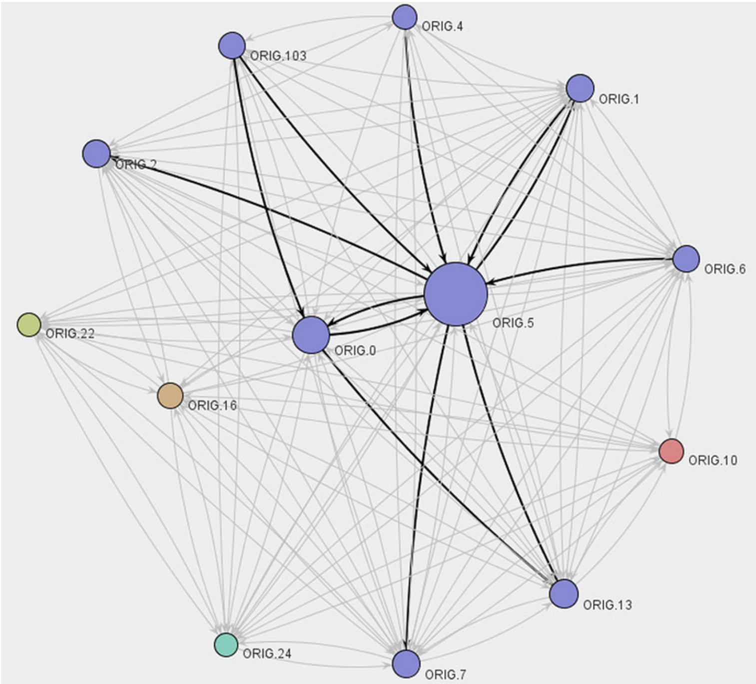


Fig. 13.10 Social network consisting of civil servants that executed more than 2000 activities in a 9 month period. The darker arcs indicate the strongest relationships in the social network. Nodes having the same color belong to the same clique. Names of resources have been anonymized for privacy reasons

ered for another municipality (i.e., different from the one for which we analyzed the WMO process). We applied the heuristic miner on an event log containing information about 745 objections against the so-called WOZ ("Waardering Onroerende Zaken") valuation. Dutch municipalities need to estimate the value of houses and apartments. The WOZ value is used as a basis for determining the real-estate property tax. The higher the WOZ value, the more tax the owner needs to pay. Therefore, Dutch municipalities need to handle many objections (i.e., appeals) of citizens that assert that the WOZ value is too high. For this municipality we analyzed four processes related to objections and building permits. Here, we restrict ourselves to the WOZ process shown in Fig. 13.11.

The discovered WF-net has a good fitness: 628 of the 745 cases can be replayed without encountering any problems. The fitness of the model and log is 0.98876214 indicating that almost all recorded events are explained by the model. Hence, the WOZ process is clearly a Lasagna process. Nevertheless, it is interesting for the municipality to see the deviations highlighted in the model. Figure 13.12 shows a fragment of the diagnostics provided by the conformance checker (cf. Sect. 8.2).

The average flow time is approx. 178 days. Figure 13.13 shows some more performance-related diagnostics computed while replaying the event log contain-

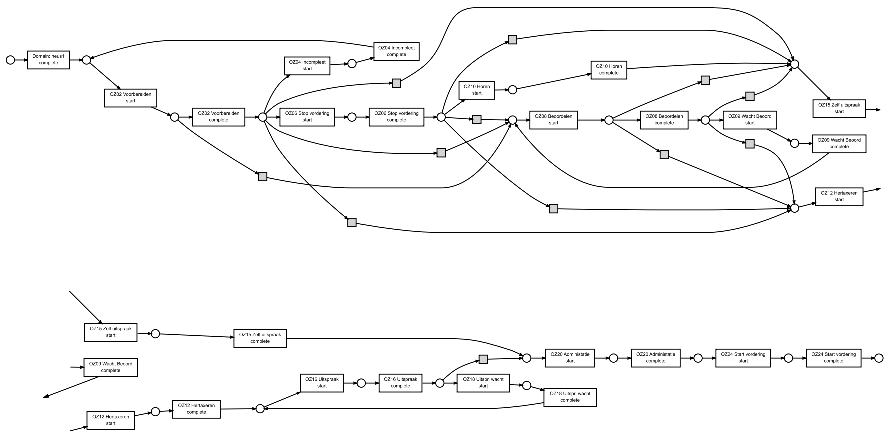


Fig. 13.11 WF-net discovered based on an event log of another municipality. The log contains events related to 745 objections against the so-called WOZ valuation. These 745 objections generated 9583 events. There are 13 activities. For 12 of these activities both start and complete events are recorded. Hence, the WF-net has 25 transitions

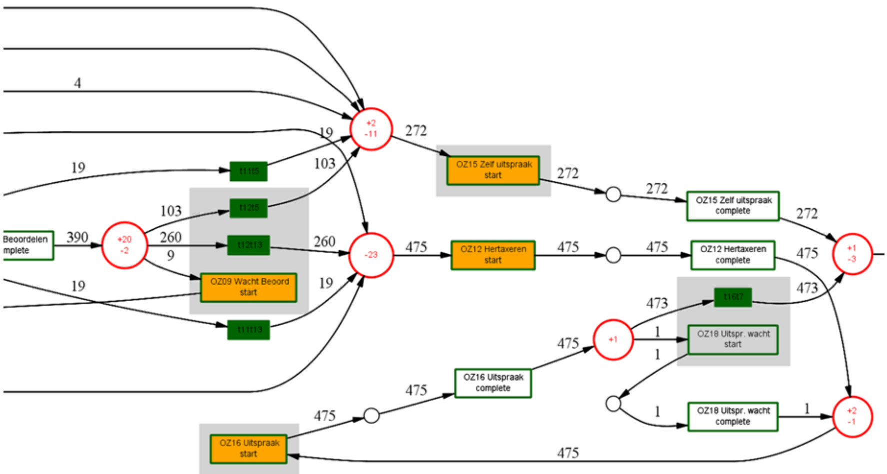


Fig. 13.12 Fragment of the WF-net annotated with diagnostics generated by ProM's conformance checker. The WF-net and event log fit well (fitness is 0.98876214). Nevertheless, several low-frequent deviations are discovered. For example, activity "OZ12 Hertaxeren" (re-evaluation of WOZ value) is started 23 times without being enabled according to the model

ing timestamps. The standard deviation is approx. 53 days. ProM also visualizes the bottlenecks by coloring the places in the WF-net. Tokens tend to reside longest in the purple places. For example, the place in-between "OZ16 Uitspraak start" and "OZ16 Uitspraak complete" was visited 436 times. The average time spent in this place is 7.84 days. This indicates that activity "OZ16 Uitspraak" (final judgment)

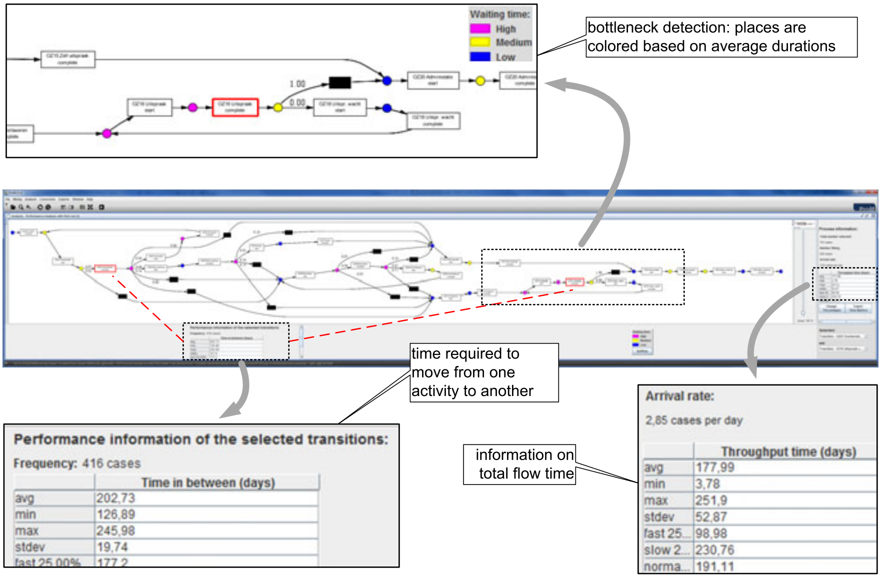


Fig. 13.13 Some diagnostics obtained by replaying the event log. These diagnostics explain why objections take on average approx. 178 days to be handled

takes about a week. The place before "OZ16 Uitspraak start" is also colored purple; on average it takes 138 days to start this activity after enabling. As shown in Fig. 13.13, it is also possible to simply select two activities and measure the time that passes in-between these activities. On average 202.73 days pass in-between the completion of activity "OZ02 Voorbereiden" (preparation) and the completion of "OZ16 Uitspraak" (final judgment). Note that this is longer than the average overall flow time. This is explained by the observation that only 416 of the objections (approx. 56%) follow this route; the other cases follow the branch "OZ15 Zelf uitspraak" which, on average, takes less time.

The event log also contains information about resources. The 9583 events are executed by 20 resources. Most activity instances have a start and complete event. These are typically done by the same person. However, in exceptional situations an activity is started by one person and completed by another. Table 13.1 shows the resource-activity matrix introduced in Sect. 9.3. The table shows that some people executed many activities (e.g., user 8 generated 2621 events) whereas others executed just a few activities (e.g., users 13 and 14 generated only one event). Figure 13.14 shows a social network based on the user profiles shown in Table 13.1. Persons that have similar profiles are connected and the strength of a connection depends on the degree of similarity (here we used the correlation coefficient). This information can be used to group people. Figure 13.14 shows four cliques discovered by ProM's social network analyzer: *clique 1* consists of users 1, 2, 3, 8, 12, 13, 14, 16, and 17, *clique 2* consists of users 4, 5, 6, 9, 11, 18, and 19, *clique 3* consists of users 7 and 15, and *clique 4* consists of users 10 and 20. Consider, for example,

Table 13.1 Resource-activity matrix showing the number of times each user performed a particular activity: a_1 = "Domain: heus1", a_2 = "OZ02 Voorbereiden", a_3 = "OZ04 Incompleet", a_4 = "OZ06 Stop vordering", a_5 = "OZ08 Beoordelen", a_6 = "OZ09 Wacht Beoord", a_7 = "OZ10 Horen", a_8 = "OZ12 Hertaxeren", a_9 = "OZ15 Zelf uitspraak", a_{10} = "OZ16 Uitspraak", a_{11} = "OZ18 Uitspr. wacht", a_{12} = "OZ20 Administatie", a_{13} = "OZ24 Start vordering". The names of users have been anonymized for privacy reasons

User	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
User 1	0	0	51	0	0	0	0	0	0	0	0	0	0
User 2	1	2	0	0	2	0	0	0	0	38	0	69	0
User 3	0	9	0	0	0	0	0	0	0	0	0	0	0
User 4	2	0	0	0	0	0	0	0	0	0	0	0	0
User 5	117	0	4	0	3	0	0	0	0	1	0	20	6
User 6	172	6	14	0	7	3	0	0	1	2	0	48	53
User 7	1	41	8	14	275	8	8	865	55	180	0	128	5
User 8	2	868	7	6	105	0	0	79	266	441	0	844	3
User 9	90	0	2	0	1	2	0	0	1	2	0	27	28
User 10	0	0	0	899	0	0	0	0	0	0	0	0	1019
User 11	336	1	3	1	4	2	0	0	0	1	0	18	23
User 12	1	645	13	21	419	3	0	3	217	281	1	334	9
User 13	0	1	0	0	0	0	0	0	0	0	0	0	0
User 14	0	0	0	0	0	0	0	0	0	1	0	0	0
User 15	0	0	0	0	0	0	0	2	2	0	0	2	0
User 16	1	3	3	2	1	0	0	1	2	3	1	0	0
User 17	0	4	0	0	0	0	0	0	0	0	0	0	0
User 18	9	0	0	0	0	0	0	0	0	0	0	0	0
User 19	13	1	0	0	1	0	0	0	0	0	0	4	0
User 20	0	0	0	21	0	0	0	0	0	0	0	0	258

clique 4. The two persons in this clique (users 10 and 20) only execute a_4 ("OZ06 Stop vordering") and a_{13} ("OZ24 Start vordering"). Hence, it makes perfect sense that they are grouped together. For organizations it is interesting to see whether such clusters correspond to existing roles. Unexpected outcomes may trigger a redistribution of work.

The municipality for which we analyzed the WOZ process, provided us with several other event logs. For instance, event logs related to the handling of building permits. All of these processes can be classified as Lasagna processes and in principle all of the process mining techniques discussed in this book can be applied. The application of conformance checking on the processes of this municipality is discussed in more detail in [121]. For example, there it is shown that, despite the presence of a WFM system, processes still deviate from the normative models. The municipality was using eiStream WFM system (formerly known as Eastman Software and today named Global 360), therefore, we did not expect any deviations. However, as discussed in [121], process mining could reveal misconfigurations of the WFM

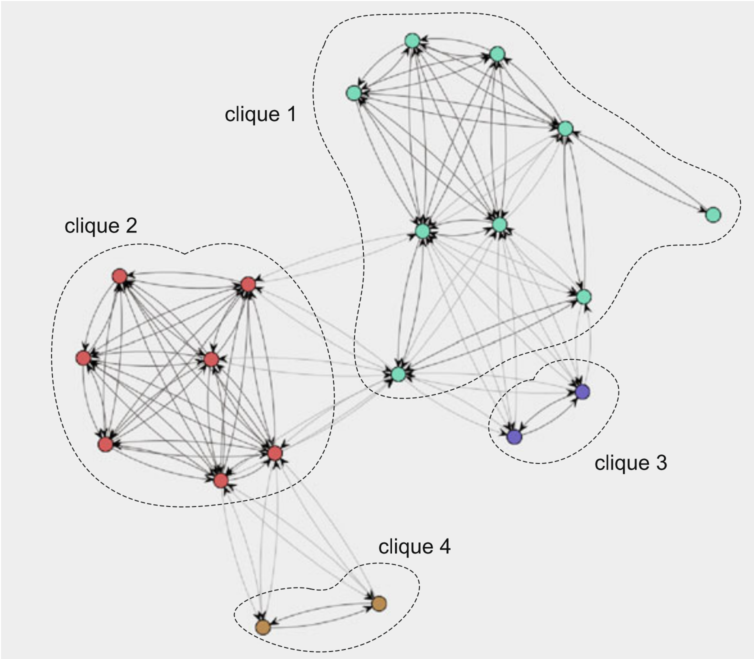


Fig. 13.14 Social network based on similarity of profiles. People that execute similar collections of activities are related and clustered in cliques

system. In [124], it is shown that, based on the event logs of this municipality, it is possible to discover simulation models covering all perspectives (control-flow, data dependencies, performance characteristics, and organizational characteristics). In Sect. 9.6, we showed how these perspectives can be merged into a single CPN model that can be simulated by CPN Tools. Although we did not conduct short-term simulations for this municipality, the validation of the models described in [124] shows that accurate simulations are possible for the selected process. Similarly, we showed in [167] that accurate time predictions are possible for the WOZ process of this municipality. In [167], various annotated transition systems are constructed using the approach described in Sect. 10.4. Each of these annotated transition systems is learned using one half of the event log, and evaluated using the other half. This illustrates that operational support is indeed possible for Lasagna processes.