algorithms

Article

An Experimental Outlook on Quality Metrics for Process Modelling: A Systematic Review and Meta Analysis

Ashish T. S. Ireddy and **Sergey V. Kovalchuk** *

Faculty of Digital Transformation, ITMO University, Saint Petersburg 199034, Russia; ireddy@itmo.ru
* Correspondence: kovalchuk@itmo.ru

Abstract: The ideology behind process modelling is to visualise lengthy event logs into simple representations interpretable to the end user. Classifying process models as simple or complex is based on criteria that evaluate attributes of models and quantify them on a scale. These metrics measure various characteristics of process models and describe their qualities. Over the years, vast amounts of metrics have been proposed in the community, making it difficult to find and select the appropriate ones for implementation. This paper presents a state-of-the-art meta-review that lists and summarises all the evaluation metrics proposed to date. We have studied the behaviour of the four most widely used metrics in process mining with an experiment. Further, we have used seven healthcare domain datasets of varying natures to analyse the behaviour of these metrics under different threshold conditions. Our work aims to propose and demonstrate the capabilities to use our selected metrics as a standard of measurement for the process mining domain.

Keywords: process modelling; metrics; evaluation; quality; business process modelling

check for updates

Citation: Ireddy, A.T.S.; Kovalchuk, S.V. An Experimental Outlook on Quality Metrics for Process Modelling: A Systematic Review and Meta Analysis. *Algorithms* **2023**, *16*, 295. https://doi.org/10.3390/a16060295

Academic Editors: Sally McClean, Adele Marshall, Ian McChesney and Zeeshan Tariq

Received: 1 May 2023
Revised: 4 June 2023
Accepted: 7 June 2023
Published: 10 June 2023

1. Introduction

Process mining has advanced rapidly within the past two decades and is presently shaping itself into a foundational domain. Since the coining of the term *"Process Mining"*, the amount of research and evolution that has taken place within is vast. These spheres of expansion are majorly classifiable into topics such as algorithmic approaches, visualisations, modularity, optimisation, data interpretation and analysis. Together, the progress from these spheres gave rise to the commercialisation and expansion of the domain from a theoretical concept into a practical reality. Experimental articles on data discovery algorithms and mining concepts have transformed into software toolkits and marketable products used in the field by organizations to track and improve their processes. The years leading up to 2010 depict that most research efforts were focused on the conceptual part of analysis aimed at creating new mining algorithms, improving visualisations and other core elements utilised in process modelling [1–6]. Further on, the focus shifted towards enhancement and evaluation of generated models [7–10]. This transition was in sync with the presence of better technology. Upon achieving progress, process modelling was employed in many fields at various levels of application (i.e. from basic procedural descriptions to supply chain routines). This brought along the issue of complexity. Excess or lack of information in the model proved to be a failure of process mining as a technique. The only remaining solutions is to analyse the model and determine if it is fit for the end user. Hence, by evaluating and measuring a process model, insight into the level of interpretability and the technical viability of the data is revealed. The cornerstone is being able *"To make sure we achieve the objective"* (i.e. in principle, a better understanding of our data).

In the pipeline of process modelling, the stages are generally in the order of data collection, application of algorithms, visualisation and evaluation. Our work aims to propose a solution on the part of process model evaluation. The main motivation for our study was derived from a situation where we wanted to measure a model. To do so we had to learn the fundamentals of process model assessment and comb through multiple articles to find a matching metric for our data. Further, we had to interpret the meaning of the metric in relation to a scale and continue searching until the right fit was found. This was a tedious and time-consuming task.

When trying to find suitable measures, we encountered a never-ending supply of metrics described within individual works and a few summarised in review works. We did not find a clear solution such as an ISO standard that recommends the use of certain techniques or measures for the whole sector. Further, we observed that many metrics proposed by researchers in the community have both unique and universal applicability. This implies that every metric has its specific use and cannot be overlooked. However, from the perspective of universal modelling, we did not find a concise set of metrics that can be applied to all process models with universal attributes. Our work aims to address these issues by presenting a definitive solution to approach process model evaluation and metric-based assessment. Thus, our research objectives were crafted as follows:

- OBJ 1: To propose a concise solution to the vast amount of metrics in the scientific community;
- OBJ 2: To demonstrate the use of metrics aimed to showcase the evaluation of process models.

We have performed two individual literature reviews that provide insight into process model evaluation and its history. A taxonomy presents a comprehensive solution to the vast amount of metrics proposed in the community. In our work, we have focused on the technical aspect of model evaluation and have created a classification of metrics based on human-centric qualities of *understandability*, *complexity* and *interpretability*. Lastly, we have summarised the four most widely used metrics in the sector and have benchmarked their performance using real-world healthcare event logs. We have analysed their behaviours under various control conditions and have proposed to use them as a standard of measurement within process modelling.

Further, the paper is divided into the following sections: Section 2 consists of the background and the literature survey. Section 3 elaborates on the methodology, implementation and results of our benchmarking experiment. Section 4 consists of an open discussion of various thoughts, ideas and questions faced over the course of this work. Section 5 is the conclusion. Appendixes A–C consists of material that can be used by the community for future research.

2. Literature

2.1. Background

As mentioned earlier in Section 1, process modelling has been flourishing since the early 2000s. Today, it has become a commercialised sector and is further developing into its own field. During this phase of transitioning into a sector, there came a stage when we needed to assess the fruition of results. This is where the evaluation of process models is primarily required. Much research has been carried out and numerous metrics have been proposed, yet there still exists an uncertainty on how to perfectly evaluate process models. The measurement of models is commonly performed using their characteristic attributes, but as the volume and variety of data grew, many secondary features were being introduced for modelling. By utilising non-universal features for assessment, newly developed metrics were restricted to the type of data that can be used for measurement. Therefore, there is a need to distinguish and classify metrics that can be used on models generated from universal event logs. In our work, we have prioritised selecting metrics and models that use universal elements for assessment and generation.

Today, we have reached a point where metric evaluation can produce a numeric value of sorts that describes the model. However, the details of this rating are left to the user to decode (i.e., the meaning of the metric itself and its relational rating of the process model on that respective metric scale). This creates a complex environment for approaching metrics. Further, the volume of metrics already proposed in the community makes it difficult to

specifically choose a metric without overlooking another. Process model qualities were created to address this issue. They were used to distinguish certain properties of the model that can be interpreted by the user via metrics. In our work, we have chosen three major qualities crucially relevant to describe the technical properties and the human perception of the model.
We collectively address these issues using three crafted research questions. By answering these questions, we were able to achieve our research objectives:

- RQ1: What are the methods to evaluate a process model?;
- RQ2: How has process model evaluation progressed over the years?;
- RQ3: How do we measure the understandability, complexity and interpretability of process models using metrics?

Research questions RQ1, RQ2 and RQ3 are addressed as follows. ***RQ1*** and ***RQ2*** are answered using a broad spectrum study along with a meta-review. The broad review gives a ground-up perspective of model measurement and provides a simplified outlook on its evaluation and assessment. The meta-review is a state-of-the-art *review of reviews*. It presents a taxonomy of established metrics proposed by the community and aims to list all the viable metrics within a tabulation. ***RQ3*** is answered in Section 2 using a diagram and via an experiment in Section 3. This answers the question of evaluating process models based on their *understandability, complexity and interpretability*. The experiment demonstrating the usage of evaluatory metrics on healthcare event logs is also described in Section 3. We have filtered down to four prime metrics that operate based on the fundamental principles of process modelling (i.e., using universal features). The latter objective of *having uniform metrics for the process mining domain* as a whole is answered using the results generated from our experiment. Our work further aims to provide *"A once and for all solution"* to the vast amount of metrics in the community and propose a standard of measurement for process modelling.

2.2. What Are the Methods to Evaluate a Process Model?

This research question is answered using a broad-spectrum study. It aims to give a general idea of metric definitions, conditions and criteria required to qualify as a measure.

Search and Approach: The broad spectrum study utilises a rebounding fashion of analysis. We have navigated through the area of model measurement, contrasting various information sources and aiming to obtain an overview of how to evaluate process models. The literature search for this study was relatively mainstream. Figure 1 provides an overview of the search methodology. For this review, the idea was to present wide-ranging information and not dive into specific topics. Hence, generalised search terms were selected to include wide-ranging topics of modelling and evaluation. We considered the risk of bias factor when creating the keyword search strategy and addressed this aspect by using wider terms therefore almost nullifying the tendency to tilt towards a single area.

We used Google Scholar's advanced search options to find research works as it allows for a bounded exploration of the literature. It was set up as follows: the articles with *"the exact phrase"* occurring *"anywhere in the article"* between the dates 2010 and 2022, with the remaining fields left blank. The generation and usage of keywords were carried out in two different ways. The first approach involved using individual keywords such as ***"Process models"***, ***"Business process models"***, ***"Process mining"*** and ***"Process modelling"*** as singular terms in the field. The second approach employed the use of multiple terms, where the first word ***"Process Model"*** remained constant and singular words such as ***"Metrics"***, ***"Evaluation"***, ***"Quality"***, ***"Measurement"***, ***"Complexity"***, ***"Interpretability"***, ***"Quality dimensions"***, ***"Quality metrics"*** were substituted in place of each other (i.e., "Process Model" "Metrics"). Similarly, a second search was conducted using the second approach with the following terms: ***"Process Modelling"*** and ***"Framework"***, ***"Guidelines"***, ***"Rules"***. From the above search queries, 624 articles were identified to be within the scope of the broad spectrum review. This number of articles is massive. To solve this, we crafted generalised queries and utilised them as pillars of reference when reading the full text. ***Query 1:*** *Understanding metrics for process models*—the query directs focus towards studying the literature that defines metrics for process models. ***Query 2:*** *Perceiving methods and approaches to evaluate process models*—this query focuses on analysing works that propose and demonstrate the usage of metrics. ***Query 3:*** *Scrutinising criterion to qualify as a metric*—this query concentrates on works that highlight rules, conditions and criteria for metrics. The selected works were classified into five groups. We cross-referenced between these sources to obtain definitive solutions to our questions.

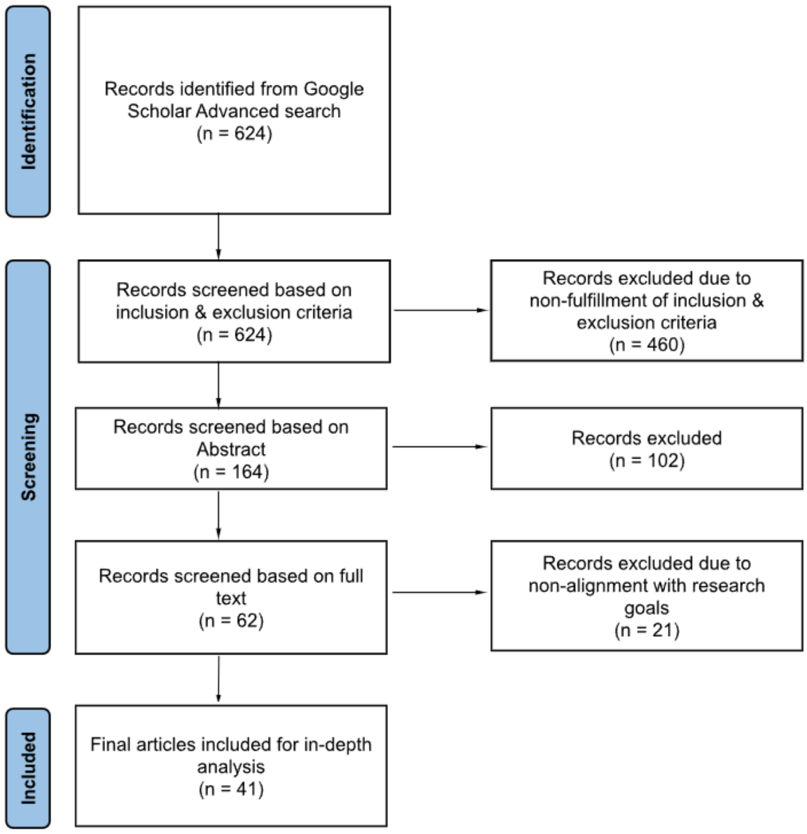


Figure 1. Selection procedure for broad-spectrum study.

1. ***Process Modelling Books***: Works listed in this group consist of books written to cover prime directions and overview of the field [11–17];
2. ***Existing Metric Articles***: This group consists of works that present metrics and quality dimensions to the community [18–26];
3. ***Frameworks***: This group consists of articles that showcase the quality metric framework specifically used to address the classification of metrics and their characteristics [9,27,28];
4. ***Guidelines***: These articles consist of publications that emphasise and describe conditions required for quality dimensions to qualify as a metric [29–34];
5. ***Evaluatory Applications***: These works present exemplar applications that demonstrate the usage of metrics and quality dimensions on actual data [21,35–42].

Results of the Broad Spectrum Review:

Query 1: *Understanding metrics for process models.* Within process modelling, the terminology of evaluation is broad. Evaluation, metrics and quality dimension are a few terms used to showcase the variation in features of a process model. To obtain a clear understanding of the definition of a metric we looked towards the ISO org [43], which says that a metric is a quantifiable measure used to assess the performance of an object relative to a scale. In process modelling, process graphs are the main objects of measurement generated from the data. Likewise, we should aim to evaluate the model and not the event log. It was determined that any generated model has to satisfy a specific set of requirements and criteria to classify as fit-for-purpose. A model can be considered of high quality only when it is fit-for-purpose. Utilising global characteristics from the event log and the process graphs yields insight into the model's viability for purpose [44–47]. Additionally, quality dimensions can be used to measure traits of the model. These dimensions are differentiated based on their constituting attributes from the process model and the resulting quality shown as in the CPMQF framework [48]. **Query 2:** *Perceiving methods and approaches to evaluate a process model.* The query is answered by splitting the analysis into two major categories. *Syntax-based evaluation*: This category involves using attributes of the process model that are directly extractable and do not require additional features for calculation (e.g., lines of code, number of nodes, number of elements, number of connections, etc.) *Characteristics-based evaluation*: These are defined terminologies that exhibit behaviour extracted from the process model (e.g., understandability, simplicity, modifiability, comparability, complexity etc.) **Query 3:** *Scrutinising criterion to qualify as a metric.* This query elaborates the conditions for a dimension to be termed a metric. To understand how metrics were established and accepted, we have turned towards works that certified and created criteria for researchers to propose their evaluatory metrics. It involved articles including [29–32], where there are specific conditions that a metric has to satisfy to qualify as a quality measure. The ideal information for query 3 is present in frameworks, guidelines and standards. The following are five of the most widely used guidelines for process modelling metrics:

1. *Guidelines for business process modelling (2000) (GoM)* [49]: This is one of the first foundational guidelines and frameworks developed to create an approach to measure the qualities of process models;
2. *The 7PMG (2010)* [30]: A revised and pinpointed version of conditions that a quality dimension or a measure should aim to fulfill in order to generate a process model of good quality;
3. *Pragmatic guidelines for business process modelling (2014)* [29]: A detailed approach to understand the logical and model generation part of quality defintion;
4. *Quality assessment strategy (2017)* [50]: A quality assessment strategy proposed to create evaluation criteria to assess new models;
5. *Process modelling guidelines (2018)* [31]: This is a systematic literature review and experiment showcasing the various metrics, summarised using an internal relevance.

Other frameworks and regulations mentioned along with these works are CPMQF [48], SEQUAL [51,52], CMQF [53], GoM [49], SIQ [54], BPMQ [55] Frameworks and the ISO standard (ISO 25010:2011) [43].

Overall, from the broad spectrum study, it was found that process models are evaluated based on particular attributes. These attributes vary between the syntax features of process models and properties of the event log itself. According to the search, the ideal way to evaluate a process model is to perform calculations between the model and its respective event log. These calculations are founded upon the ***Basic Three*** features (i.e., *Unique ID, Activity and Time Stamp*) [56]. Further, secondary attributes can be used for extended analysis of the process model and the event log, but with customised methods of calculation.

2.3. Meta Review and Evolution

The research questions RQ2 and RQ3 are addressed within this section. The meta-review provides a collective overview of process model measurement over the years and summarises the vast amount of metrics using a taxonomy. The practical part of RQ3 is extended into an experiment in Section 3.

Early approaches for process model evaluation were solely based on software engineering techniques. However, standalone tools and software (e.g., ProFIT [57], PM4PY [58] etc), are replacing these techniques. The current generation of metrics is established on the three basic attributes. In our work, we have focused on the quality dimensions of **understandability**, **complexity** and **interpretability** [59–62]. We believe these qualities provide a balanced outlook on human assessment of the models. *The level of easiness in perceiving a model is called understandability. The amount of detail shown via the process model from the event log is called complexity. A combination of both complexity and understandability is termed as interpretability* [63,64]. When trying to evaluate models, it seemed as though there was a non-stop supply of research works aimed at process model evaluation and quality analysis. Hence, a clear representation of viable metrics for process modelling is required. Even after obtaining an understanding and overview of the various metrics present in the community, there remains a small gap. Implementation and demonstration, it follows through by providing results to the community, which will further aid researchers in shaping the evaluatory phase of process modelling. We have covered the demo of our selected metrics in Section 3. The meta-review aims to answer the follow-up questions from RQ1 and create a taxonomy of metrics. Our search approach for the meta-review was rather unique. The results from our broad spectrum review were taken forward and used to create a search procedure that is more efficient than manual searches.

The meta-review is a *review of reviews*. The goal is to create a taxonomy of all the metrics proposed within the field of process modelling. The taxonomy can be considered a crucial element for researchers who aim to work in conformance checking and evaluation of models, as it provides an explicit representation of the various metrics present. To create this taxonomy, literature reviews published through the years that summarise the most sought-after metrics were used as critical sources. The taxonomy is included in Appendix C. It lists all the metrics proposed between 1997 and 2023. An individual study of metrics was not performed, since it is tedious and might leave us with a bulk of information with broad relevance. Other issues such as metrics with different names but the same working principle, metrics with very distinct methods of evaluation (i.e., using non-universal features that might not be available in all datasets) and metrics with lengthy calculations or unverifiable performance were considered as possible instances that would exponentially increase the time and energy required to perform a study. Hence, review articles were selected as the source of metrics. As most review works explicitly mention the source of metrics, it makes them efficient sources to analyse. A motto was kept in mind when creating the meta-review, "*i.e., to make sure that the proposed information will be useful to the community and not be an additional study weight*".

Search Procedure: The literature search for the meta-review was far-reaching and compact. As our strategy was to use review articles, we were able to minimise the depth of each search term in exchange for a wider array of topics. This, in turn, allowed us to precisely find review articles that are in the sphere of process model assessment and not singular works. Similar to the broad spectrum review, we considered the risk of bias factor when creating the keyword search strategy; this resulted in an increased number of search terms. The search terms were generated based on the spheres that deal with process model evaluation, metrics, quality, modularity and assessment. Since process mining has many neighbouring fields that share the same form of evaluation (e.g., business process modelling, workflow), we specifically selected terms to avoid any bias or overlooking any research works. The primary search was performed using Google Scholar search and the secondary search was conducted through Scopus. The search using Scopus was mainly to recover any missing articles that were not found during the primary search via Google Scholar. Similar to that of the broad spectrum review, the Google Scholar advanced search option was set up in two ways.

The first approach was set up as follows. The articles with *"the exact phrase"* occurring *"anywhere in the article"* between the dates 1997 and 2023 with the *Review articles only* field enabled and the remaining fields were left blank. The keywords used for this search were ***"Workflow learning", "Workflow Quality", "Workflow discovery", "Workflow Discovery Quality", "Workflow Discovery Evaluation", "Business Process Modelling", "Business Process Modelling Metrics", "Business Process Modelling Quality", "Business Process Model Quality", "Business Process model quality indicators", "Process Learning", "Process Mining", "Process Mining Metrics", "Process Modelling", "Process Model Notation",***

"Process Model Quality", "Process Model Evaluation Quality", "Process Model Understandability", "Process Model Complexity", "Process Model Understandability", "Process Modelling Metrics", "Knowledge discovery workflow".

The second search approach was set up as follows. The articles with *"with all of the words"* and *"with the exact phrase"* occurring *"anywhere in the article"* between the dates 1997 and 2023 with the *review articles only* field enabled and the remaining fields were left blank. The keywords were used as follows. The first part of the phrase was entered in the first field of *"with all of the words"* (i.e., the phrase before AND) and the second part of the phrase was entered in the field *"with the exact phrase"* (i.e., the phrase after AND). The search terms were **(*"Quality assessment"*** AND ***"Business process models")***, ***"(Quality indicators"*** AND ***"Business process models)"***, ***"(Process model complexity"*** AND ***"Metrics)"***, ***"(Process learning"*** AND ***"Process mining)"***, ***"(Process discovery"*** AND ***"Process mining)"***, ***"(Modularity representation"*** AND ***"Business process models)"***, ***"(Presentation medium"*** AND ***"Business process models)"***, ***"(Experimental evaluation"*** AND ***"Business process modelling)"***, ***"(Business process model"*** AND ***"Quality)"***, ***"(Business process model"*** AND ***"Complexity)"***.

The articles obtained from the search terms were examined by their abstract and full text to make sure they did not fall into neighbouring fields such as networking and business informatics. Figure 2 exhibits the search procedure for the meta-review. The following inclusion and exclusion criteria were used to filter works:

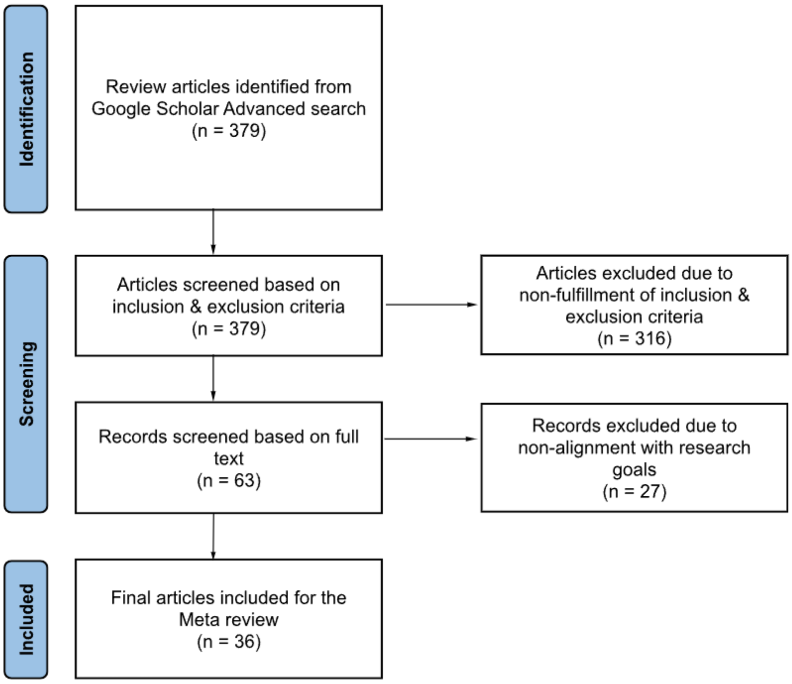


Figure 2. The selection procedure for the meta-review.

Inclusion criteria:

- Should be a review paper;
- Should have been published between 1997 and 2023 (when performing this study, articles published until January 2023 were considered for analysis);
- Should be relevant to process mining and business process modelling;
- Should be aimed at quality, measurement, modularity, evaluation, metrics and complexity for process models or modelling.

Exclusion criteria:

- Works that focus on topics other than evaluation (e.g., process algorithms, languages, tools or mining approaches, visualisations, oncology);
- Works that focus only on one metric or one characteristic;
- Works not in English.

Results of the Meta Review: We found 379 review articles published in the time frame between 1997 to 2023. Each work summarises various evaluation methods and metrics for process modelling. Upon filtering works based on inclusion and exclusion criteria, a total of 63 articles were obtained. After reading the full text, 36 articles were eligible for analysis and inclusion in the meta-review [65–70]. We listed 140 metrics proposed by fellow authors within our meta-review. They are categorised based on their impact on the model qualities of understandability, complexity or interpretability. Additionally, we have included new fields that would aid researchers in selecting and implementing proposed metrics. The taxonomy of metrics consists of the following fields of data (the respective acronyms are included in parentheses):

1. The **Name** of the metric (Metric);
2. The **Year** of proposal (Year);
3. **Authors** of the metric (Auth);
4. Working **Definition** (Def.);
5. Status of **Demonstration** by the authors *(Yes/No)* (Demo):
    - Yes (Y)—A demonstration of the metric has been provided by the authors in their works;
    - No (N)—A demonstration of the metric has not been provided by the authors;
6. Type of **Metric** (Type):
    - *Scratch (S)*—Metrics that are completely designed by the author(s);
    - *Partially Derived (PD)*—Metrics that borrow ideologies but have their own mechanism;
    - *Derived (D)*—Metrics implemented based on existing concepts and measures;
7. Source to **Original paper** (Src).

Evaluation History: Early literature collectively describes metrics as one of the "*goals of models*" [71]. We have come up with a simple representation of metrics in retrospect to process mining as a whole. Figure 3 represents the viewpoint of evaluation and its branching. Level 1 consists of the main branches of the field and splits into process discovery, knowledge mining and conformance checking. Level 2 consists of topics which come under each level 1 division. These topics are subdivisions that focus on specific aspects of evaluating models. *Knowledge mining* consists of methods used to extract information from logs and datasets. *Conformance checking* encompasses approaches to analyse the mining performance while discovering the event logs and extracting information. The *process discovery* division deals with performing visualisations and implementation of algorithms to generate models. We believe that these subdivisions work hand-in-hand when dealing with evaluation. Level 3 is a combination of L1 and L2. It ideally performs the model quality analysis. Level 3 is classified as quality dimensions, evaluation and metrics.

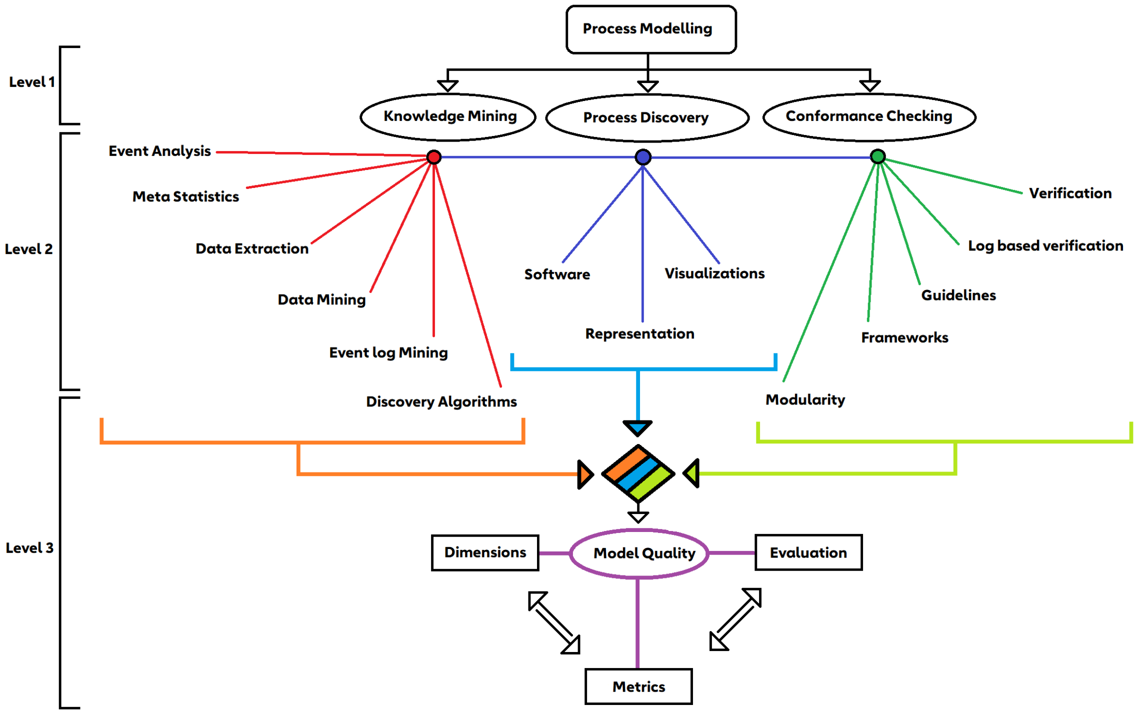


Figure 3. Branching of evaluation in process modelling.

The goal of evaluation is to have metrics that can measure specific qualities of a process model. Hence, we have focused on model qualities that aid in improving the human perception of the model. We believe that understandability, complexity and interpretability can be coined as crucial qualities for human perception of process models [72]. To ensure that our taxonomy remains simple and allows for easy traceability to the original works, we adapted strategies to enhance the results:

- Ensure the metrics listed were within the scope of modelling qualities: understandability, complexity and interpretability;
- Link metrics that were conceptually the same but with different authors and explicitly mention them;
- Perform a viability check of the measures (metrics listed should be ideal for universal application, i.e., should utilize the basic three);
- Scrutinise and check if the proposed metric is demonstrated by the author and can be replicated by prospective researchers;
- Mention redundant metrics where necessary;
- Include a mini description to improve readability;
- List the original source of the metric.

The work by [4] in 1997 exhibits the necessity of evaluation in business process modelling. The authors emphasise representing factual data in a simplified manner (i.e., to improve the human perception of the data). With this in mind, if a generated process model does not provide a certain level of data interpretability, then the modelling procedure may be termed a failure. Authors of [73], in 1998, coined the term direct flows graph. It is a more flexible and less accessory-oriented representation for process modelling. The use of DFG has simplified the learning curve to approach process modelling. It has also aided in minimising the gap between data modellers and the end user. Unlike Petri nets and process graphs, the DFG shows only required information such as flow, activities and other peripheral details. The logical section is hidden under a layer and need not be shown along with the final resultant. Using DFGs as a tool was scarce during the early 2000s, as it was a fresh concept still in adaptation.

The boom in technology played its part in the rapid digitisation of the world. It gave rise to enormous amounts of process data usable for analysis and experimentation. As data increases, so does the complexity of the process model. Thus, the trend started from here; to resolve the complexity of modern processes, we need a way to distinguish simple and complex models, i.e., easy vs. hard to interpret. During this time, process mining was also gaining a foothold in the markets as it helped analysts visualise and understand their processes much better than traditional methods [74,75].

Initially, measuring the complexity of a process model was performed by borrowing methods from neighbouring fields of software engineering and networking [76–78]. Lines of code and correctness are a few examples of metrics migrated for usage in process modelling. An increase in data generated resulted in these metrics reaching a saturation point, after which using these metrics became very tedious and cumbersome. Hence, frameworks and dimensions of measurements were proposed and developed, e.g., CPMQF [48]. Frameworks such as SEQUEL [51,52] and QME [9] (Quality Assessment framework) suggested using notations, statistics of activities, algorithm variations and many more as features to analyse the metrics of a model. From the year 2006, many works summarised the progress in the field and crafted a path to create a bridge between process mining technologies and operational business processes. The works highlighted many issues faced in the sector based on the approaches, quantifiers, analysis of results, representation and differences within. Nevertheless, the principle remained the same: ensuring process information was conveyed easily using process models [79–82].

Parallelly, the implementation of new algorithms took place with enhanced feedback. The results focused on improving the algorithms and not just the visualisations. Ref. [83] lists the most-ideal ways to judge the complexity of the model at the time. With insight into algorithms and better feedback, modelling transformed from a static to a dynamic phase. Mining and modelling were no longer only used on past data but also for simulation and designing future usage. The usage of non-universal data attributes was high during this time. These are fields of information that might not be available in datasets of various domains. This led to a shift in using the basic three as the minimal standard for any event log analysis. The 7PMG [30] proposed seven basic guidelines for process modelling. It gave a baseline platform for evaluators and fellow modellers to generate a unified procedure for model development. The years beyond 2013 saw many researchers propose metrics linked to specific process modelling qualities (e.g., activity period, exclusiveness, generalisation, flexibility, etc.).

At the same time, Wil van der Aalst and Jan Mendling published the process mining discovery and metrics books [11,84]. Wil van der Aalst explicitly explained and started the trend of using the four prominent metrics. Other authors proposed their metrics and drove efforts to review and revise older frameworks to develop concise and present-gen versions. The progressing years saw many authors start to use a handful of metrics as a scale to compare and quantify their proposed metrics. Quality metrics including recall fitness, precision, size and generalisation were widely utilised as an informal standard.

Figure 4 shows all the metrics discovered in our meta-review; they are classified based on the three quality dimensions: understandability, complexity and interpretability. This representation of metrics by the three qualities allows researchers to select metrics that majorly read and impact a certain quality of the model. These types of diagrams can be created in the community for various process modelling qualities that list the metrics impacting them.

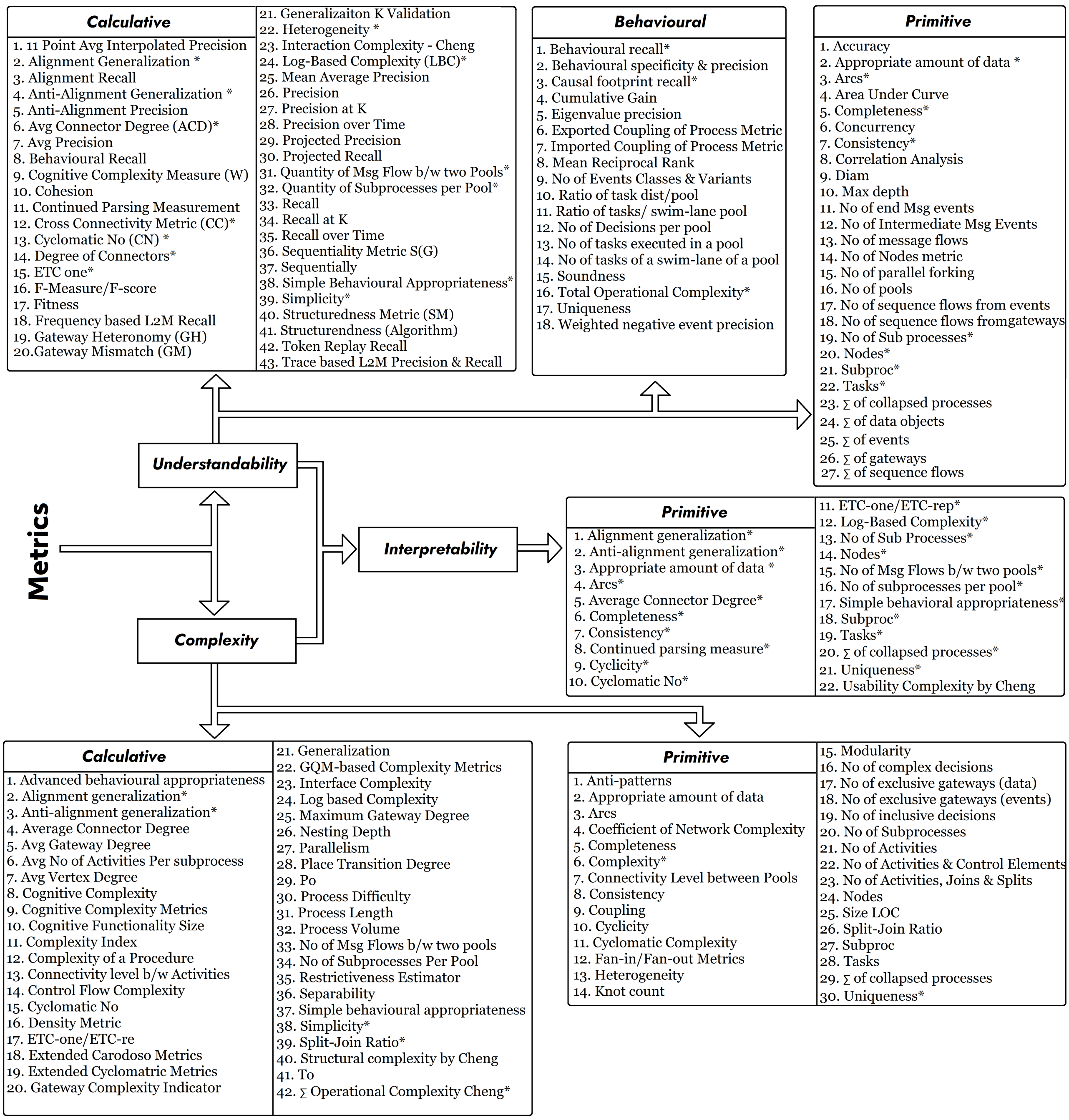


Figure 4. The taxonomy of metrics (1997–2023), (* indicates metrics that are repeating in more than one characteristic quality).

Appendix C lists all the metrics from our meta-review in detail.

We found that the metrics can be subdivided based on their technique of calculation, as follows. ***Primitive***: These are metrics that utilise older methods to be evaluated. They use brute force approaches and are evaluated directly without mathematical formulation or complex calculations. ***Calculative***: Metrics under this category require performing mathematical or algorithmic calculations to be evaluated. ***Behavioural***: These metrics are specific to the understandability quality, as they depend heavily on features of the log and require analysis of log history to be evaluated. It should be noted that some metrics are repeated within the modelling diagram and have been marked with a star (*).

One of the main motives to perform the meta-review of metrics was derived from the initial objectives: "*to rate the extent to which a process model is understandable and can be interpreted without the need for simplification*" or, in simple words, "*to be able to judge a process model based on its understandability and complexity*". Hence, in the quest to find a suitable approach, we began searching and experimenting with various metrics already published in the community. However, the thought of selecting research articles only after a certain year and ignoring the history (i.e., evolution to that point) puts us at a loss. Hence, the meta-review was conducted to give a background analysis and summary of the various approaches and techniques of evaluation to date.

3. Methodology

This section addresses RQ3 with an experiment involving real-world datasets. Using the insight from the broad spectrum study and meta-review, we found four metrics that have an ideal balance in representing the complexity and understandability of process models. We have termed them the QUAD metrics. The mathematical adaptations of these metrics were created and implemented via a process mining environment. Using real-world datasets, we have analysed the behaviour of these metrics under various control conditions. The results provide insight into how these metrics describe the technical aspect of process models.

3.1. The QUAD Metrics

We have termed ***Fitness***, ***Precision***, ***Simplicity*** and ***Generalisation*** as the QUAD metrics [85]. From the results of our meta-review, we have seen vast usage of the four metrics. They have been used as primary dimensions for evaluating process models. Initial testing revealed them to have a direct impact on the modelling qualities of understandability and complexity. We performed technical analysis and derived mathematical equations for each metric specific to our modelling environment and dataset attributes.

We have considered the following notations and terms to explain and describe the formulations of the metrics. Let E be an event log with n events and M be a process model of the same event log. T is a trace of events. C is the unique cases of events. Then, the metrics are derived as follows:

1. *Replay Fitness*: Fitness is a measure that denotes how much of the behaviour present in the event log can be reproduced by the process model. It can be elaborated as the extent to which the model can reproduce the traces recorded in the event log. We have based replay fitness of the work [86]. They have used an alignment-based approach to identify the fitness of process models. This technique aligns as many events as possible from the trace with activities in a single execution of the model. The formulation is as follows:
$$Replay\ Fitness = 1 - \frac{\mid T_E \cap T_M \mid}{\mid C_E \mid} \tag{1}$$
where, T_E is the traces from the event log, T_M is the traces from the process model, $|T_E \cap T_M|$ is the number of traces, that occur in event log and process model and C_E is the unique cases of events in the event log.
In our implementation, the numerator is calculated by aligning every case ID in the event log from start to end and observing if the model can align all of them. In simpler terms, we determine if the process model can replay every single trace as in the event log. The denominator here is the total number of unique case IDs present in the event log. On calculating, we obtain a rating between 0–1, where 1 indicates the process model can fully replay every single trace present in the event log via the generated process model and vice versa. This metric is measured on a scale of 0 to 100%, where 100% indicates the complete alignment of the traces in the event log to that of the model;
2. *Precision*: Precision is a measure used to show how clearly the process behaviour depicted in the event logs can be captured by the model without oversimplifying the model [86]. The metric aims to find out if there exist decisions that are possible in the

model but never made in the log. At first glance, it may seem rather odd, as any mining tool by definition would not intentionally add additional behaviour or, rather, what is not given as input cannot be obtained as output. However, we found that it is possible to have behaviour that is not seen in the event log since there is always a possibility of not including every combination of edges and activities within the log. The measure here is inspired by the so-called escaping edges concept [87,88], where escaping edges are representations of decisions that are not made in the log but exist only in the model. In situations when there are no escaping edges, the model is precise (state = 1). It is formulated as follows:

$$Precision = 1 - \frac{\mid T'_E \; \cap \; T_M \mid}{\mid C_E \mid} \tag{2}$$

where T_E represents the traces from the event log, T_M represents the traces from the process model, $|T'_E \cap T_M|$ is the number of traces that occur in exclusively in the process model and not the event log and C_E is the unique cases of events in the event log.
In our adaption, the numerator is calculated by checking if the process model has additional behaviour which does not exist in the log. Additional behaviour is analysed using individual traces from the event log. The denominator is calculated as the total number of traces in the event log. When new data observed in the process model is not seen in the event log, the model is considered not precise for that respective trace. It is rated between 0 and 1, where 1 signifies the model has high precision. This metric is measured on a scale of 0 to 100%, where 100% indicates no existence of unseen behaviour in the model, as per the log. There is also a contradiction with calculating precision that we have addressed in the discussions in Section 4. Another approach for precision is seen in [89];

3. *Generalisation*: Generalisation is a measure of the level of abstraction in a process model. It deals with overly precise models and tries to make sure over-abstraction is not performed. It estimates how well the process model describes an unknown system and not only the existing activities. If all parts of the process model are frequently used, then the process model is likely to be generic, i.e., generalisation is high. If parts of the model are infrequently visited then the generalisation is low. The authors [86] have based generalisation off replay fitness. They use the data obtained from fitness as a leading edge towards generalisation. The authors explain that, if a node is visited more often, then it is certain that the behaviour is more frequent and, hence, more generalised. However, if some parts of the process model are less frequently visited, then the generalisation is bad. Our adaption is based on the same concept. The formulation is as follows:

$$Generalisation = 1 - \frac{\Sigma_{a \in M_A} \sqrt{Ex(a)^{-1}}}{\mid M_A \mid} \tag{3}$$

where a denotes activities, M_A is the activities present in the model and $Ex(a)$ is the number of executions of each activity referred.
In our version of generalisation, the numerator is calculated by determining the number of times an activity in the model has been executed. The root inverse of the value is generated and a cumulative sum of all individual unique events is calculated. This metric is measured on a scale of 0 to 100%, where 100% indicates a generalised model in which all parts are used equally;

4. *Simplicity*: Simplicity is a measure that describes how easily a model can be perceived by a human subject. It can parallelly be termed model complexity. Simplicity is measured by comparing the size of the model with the number of activities in the log. Using the information from the model itself, simplicity is defined on the following principle: *if each activity is represented exactly once in the process graph, then that graph is considered to be as simple as possible* [86]. Therefore, simplicity is calculated as follows:

$$Simplicity = \frac{\mid M_D \mid + \mid E_A \; - \; MA \mid}{\mid M_A \mid + \mid E_A \mid} \tag{4}$$

where M_D is the duplicate activities present in the model, E_A is the unique activities present in the event log, M_A is the unique activities present in the process model and $E_A - M_A$ is the number of missing events in the model w.r.t to the log.
The numerator is calculated as the sum of the count of duplicate activities present in the model and the count of missing activities in the model when compared to the event log. The denominator is calculated as the sum of unique activities present in the model and the unique activities present in the event log. This metric is also measured on a scale of 0 to 100%, where 100% indicates a complex model with many events and a lower value represents a simpler model.

3.2. Experimentation

To demonstrate the use of the QUAD metrics, we used a combination of *python3* and *ProFIT*. ProFIT is a process mining tool that works based on the fuzzy heuristic miner algorithm and generates process models specifically in the direct flows graph format. The tool works similarly to other famous process mining software (PM4PY and ProM). Initially, we started the implementation of these metrics on both ProFIT and PM4PY; however, the PM4PY package did not provide the appropriate raw information to perform a ground-level implementation of metrics. Hence, ProFIT was chosen as the main tool for implementation. Currently, our implementation of metrics is specifically designed to work with input data from ProFIT; our long game is to create an open-source library (e.g., GitHub) that can be used by fellow researchers and the community to measure process models using any modelling tool. *Python3* was used to mathematically design and implement the metrics for calculation.

3.2.1. Datasets Background and Information

The datasets chosen for this analysis and experimental study of metrics were specific to the medical sphere. The selection criteria to choose the datasets are listed here:

- Should be a medical/healthcare domain dataset;
- Datasets should be open source;
- Should have sufficient events, activities and samples to generate sizeable process maps.

We searched for these datasets using Google Search and Kaggle datasets. Seven datasets fit our criteria and application. Their statistics are summarised in Table 1.

1. **Remote Patient Monitoring Data—Almazov Institute**: This dataset was provided by PMT online (an online patient healthcare monitoring system). It contains events triggered by in-home blood pressure measurements made by patients suffering from arterial hypertension [90];
2. **Hospital Records of Dutch Hospital—Eindhoven University**: This dataset was published by Eindhoven University of Technology. The dataset consists of real-life event logs of a Dutch academic hospital [91];
3. **Hospital Billing Log—Eindhoven University**: This event log was sourced from Eindhoven University of Technology. The event log was obtained from the financial modules of the ERP system of a regional hospital [92];
4. **Nurse Work Flow—Almazov Institute**: This event log was provided by Almazov National Research centre. It consists of data from the hospital access control system concerning staff activities, laboratory procedures, branch communications. etc. [93];
5. **Data Driven Process Discovery (An Artificial Event log)—Eindhoven University**: This dataset was generated by Eindhoven University. It is a synthetic event log that simulates an artificial process log of a hospital (we have used the dataset with 0% noise) [94];
6. **Central Venous Catheter Process—Conformance checking challenge 2019**: This is a dataset produced by Eindhoven University. It is also the dataset used for a Conformance checking challenge in 2019. The dataset describes the procedure to perform central venous catheter with ultrasound [95];

7. **Sepsis Treatment Pathway Dataset—Eindhoven University**: This dataset was sourced from Eindhoven University. It is a real-life event log consisting of events of sepsis cases from a hospital and its treatment [96].

Table 1. Statistics of selected healthcare datasets.

#	Dataset Name	Number of Entries	Number of Unique Traces	Average Trace length	Max Trace Size	Min Trace Size	Number of Unique Events
1	Remote Patient Monitoring Data	35,358	272	130	673	3	17
2	Real-Life Event logs	150,291	1143	131	1814	1	615
3	Hospital Billing Event Log	89,088	18,278	5	217	1	18
4	Nurse Work Flow Event log	13,644	187	73	440	1	228
5	Data Driven Process Discovery—Artificial Event log	99,589	11,112	9	11	2	8
6	Central Venous Catheter Process	697	20	35	59	26	29
7	Sepsis Treatment Careflow Dataset	15,214	26	586	8111	1	16

Upon acquiring these datasets, we explored the data using Python3 and the PM4PY and ProFIT toolkits. The discovery of data was performed in two stages:

- **Stage 1**: This stage involved sparsely overseeing the various attributes of the process model and filtering fields that may not be useful to our application (e.g., Dataset 2 had 128 fields of information, such as "special code", "diagnosis index", etc.). These additional fields were excluded and the basic three were retained in all datasets;
- **Stage 2**: In this stage, process models were visualised using Python3 and ProFIT. The Activity field in all the datasets was cleaned and made uniform with respect to all event logs. The filtering and cleaning were performed using a self-built Python script which removed all unnecessary elements in the event data (e.g., '.', '-', '_', etc.).

The two-stage filtered data were then used for all further experiments.

3.2.2. Setup

ProFIT produces a direct flows graph to visualise the process model. Direct flows graphs are a robust method that facilitate the requirements of experts and novice users. The miner in ProFIT has two control parameters to change the amount of information depicted in the process model: ***Activity rate*** and ***Path rate***. The activity rate controls the number of events (activities) shown, based on their total frequency of occurrence in the event log. The path rate controls the number of pathways (edges) connecting relative activities in the event log. Similar to the activity rate, it is controlled based on the total frequency of usage. Both rates can be varied on a scale of (0–100). In our experiment, the control parameters were varied over two regimes. These specific variations generated process models that produced a uniform pattern for analysis and generation of model ratings. The parameters were varied in the following format:
Regime 1 (R1): The *Activity rate (AR)* is varied in steps of 1 between [0 and 100] *(S1)* and the *Path rate (PR)* is varied in stages of [5], [20], [40], [60], [80], [100] *(S2)*. *(R1)* can be observed in Figure 5;
Regime 2 (R2): The *Path Rate (PR)* is varied in steps of 1 between [0 and 100] *(S1)* and the *Activity rate (AR)* is varied in stages of [5], [20], [40], [60], [80], [100] *(S2)*. *(R2)* can be observed in Figure 5.

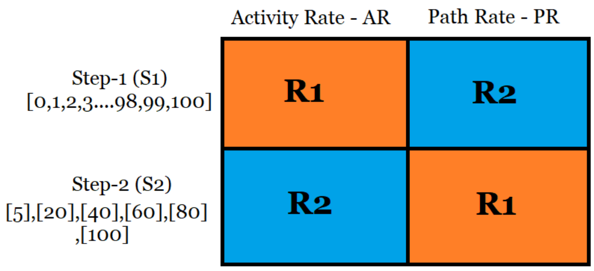


Figure 5. The variation patterns of process control features (R1 and R2).

3.2.3. Metric Calculation

Each metric was individually programmed in Python3. The metric algorithms utilise information extracted from the generated process model and respective event log. In ProFIT, the process models are rendered using graphviz. A netgraph with text-based information is parsed from ProFIT to the graphviz renderer in order to visualise the process model. We used the net graph data as model interpretation for our algorithms. The source event log remained the same for both the miner and the metric algorithms. Each metric algorithm is split into phases which perform specific calculations and produce a numeric measure at the end.

1. ***Calculation of Fitness***: Fitness is calculated according to Formula (1). The full flow of fitness calculation can be seen in Figure A1:
    - **Phase 1**: Input of event log, activity rate and path rate;
    - **Phase 2**: The event log is mined and the process model is generated. (The miner function performs filtering and cleaning of the netgraph for metric calculation);
    - **Phase 3**: Generation of a one-dimensional dictionary containing the source and destination of all activities in the model;
    - **Phase 4**: Generation of individual traces for each caseID and verification of replay via the process model. E.g., a trace is chosen from the event log and is traversed per caseID throughout the generated process model. If the trace is completely retractable from start to end, then the respective caseID is deemed fit (1), else it is deemed not fit (0);
    - **Phase 5**: Once all the caseIDs have been checked for replayability, the final formula of fitness is applied where the total number of traces that can be aligned with the model and log is contrasted against the total number of unique caseIDs in the log;
    - **Phase 6**: Generation of the percentage of fitness and output of final fitness.

For every iteration in path rate and activity rate, the algorithm is executed from phases 1 to 6. In each cycle, all caseIDs are verified for their respective fitness.

2. ***Calculation of Precision***: The precision of a model is by far the most complicated metric to be calculated, as there is a lot of ambiguity to deal with. The full flow of precision calculation can be seen in Figure A2. It is generated as follows:
    - **Phase 1**: Input of event log, activity rate and path rate;
    - **Phase 2**: The event log is mined and the process model is generated. (The miner function performs filtering and cleaning of the netgraph for metric calculation);
    - **Phase 3**: Generation of a one-dimensional dictionary containing the source and destination of all activities in the model;
    - **Phase 4**: Conversion of the one-dimensional matrix into a path matrix which uses *0s* and *1s* to indicate the existence of *"Path"* and *"No Path"* between all activities in the model. The matrix is used to generate all possible combinations from the model to verify existing and imaginary paths in the process model;
    - **Phase 5**: Generation of all possible combinations of traces in the **process model**. A permutation algorithm generates them in sets of two activities (AB, BC, AC, AA, BB, CC, etc.). Duplicates and redundancies are filtered before proceeding to phase 6;

- **Phase 6**: The trace creator, generates all possible combinations of traces from the **event log** in pairs of two (AB, BC, AC, AA, BB, CC, etc.). Duplicates and redundancies are filtered. The generated results are used to check combinations that occur only in the process model and not in the log. This value is evaluated against the total number of traces in the event log. The raw precision value of the model is generated using Formula (2). As in Fitness, for every iteration of path rate and activity rate, the algorithm is executed from phase 1 to 6. The precision is likewise calculated at every iteration;
- **Phase 7**: Generation of the percentage of precision and output of final model precision.

In phase 5, we faced an issue with the concept of precision itself, i.e., "*how many repetitions of an event can be considered for evaluation from the process model e.g., AA, AA, AA?*". We have discussed this in Section 4. In our calculation of precision, we have limited the repetitions to one (i.e., a combination can only occur once in the model).

3. ***Calculation of Generalisation***: Generalisation is calculated according to Formula (3). The full flow of generalisation calculation can be seen in Figure A3. It is calculated as follows:
    - **Phase 1**: Input of event log, activity rate and path rate;
    - **Phase 2**: Data extraction of all activities in the event log and all activities in the process model.;
    - **Phase 3**: Information from phase 2 is used to calculate the number of executions of each event in the process model and the total number of events present in the process map. These values are then used in Formula (3) to calculate the raw generalisation;
    - **Phase 4**: Generation of the percentage of generalisation and output of final model generalisation.

The generalisation is calculated for every iteration of path and activity rate and the algorithm is executed from phase 1 to 4. Individual trace extraction is not required in generalisation.

4. ***Calculation of Simplicity***: The simplicity of a model is calculated as per Fromula (4). The full flow of simplicity calculation can be seen in Figure A4. It is calculated as follows:
    - **Phase 1**: Input of event log, activity rate and path rate;
    - **Phase 2**: Data extraction of all activities in the event log and all activities in the process model;
    - **Phase 3**: Information generated from phase 2 is used to calculate The sum of all duplicate and missing activities. It is evaluated against the sum of all events in the process model and the event log;
    - **Phase 4**: The percentage of simplicity is generated for the model.

The simplicity is calculated for every iteration in activity and path rate by repeating phases 1 to 4. Similar to generalisation, individual trace extraction is not required.

3.3. Results

To understand the behaviour of metrics on the datasets, we plotted the performance of each metric in both regimes. The performance graph depicts the changes in quality ratings against one control element (either activity or path rate). We have analysed each metric individually and contrasted them against each other as groups. Every metric has two performance graphs showcasing either regime individually for all seven datasets. Within the performance graphs, each sample (S) is represented using abbreviations S1–S7.

"*Fitness is the amount of behaviour present in the event log that can be reproduced by the model*". Figures 6 and 7 show the fitness in R1 and R2, respectively. At a glance, there seems to be more than one behaviour of fitness in R1. However, by observing individually, the fitness for all datasets increases gradually or stays steady. Fitness in *S1*, *S2* and *S3* differ in steps of varying sizes. These steps indicate the addition of an event into the process model. Adding new events increases the number of paths within the model. This allows for an increased number of replayable traces. Therefore, the overall fitness of the model increases with the addition of new events. Figures A5 and A6 show an example of S1 in R1 @ PR 80. The number of paths is seen to be increasing as the number of activities increases from 9 to 12. At path rates less than 100, we can observe that the fitness does not reach 100% in any dataset. This is because there is always a certain number of paths hidden due to their low frequency of occurrence; they are only viewable at 100% control rates. Hence, the fitness does not reach 100%, as all traces cannot be replayed by the process model. In S4, S5, S6 and S7, the fitness remains constant for certain periods of activity rates as no new paths are being added to the model; therefore, there is no increase in the number of repayable traces.

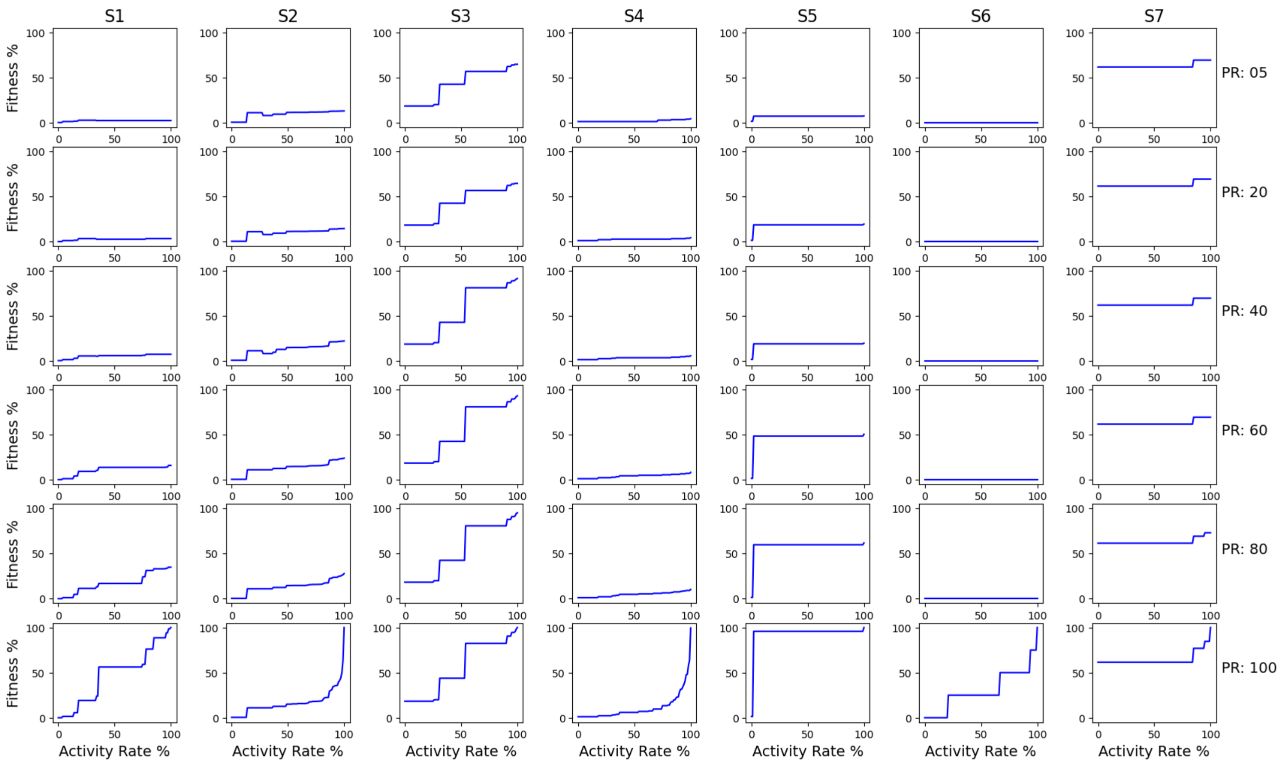


Figure 6. Fitness ratings of datasets 1 to 7 in R1.

In R2, we see similar behaviour as in R1, except the increasing nature of fitness is faster and more gradual. The number of paths seems to have more influence in replaying the traces when activities remain constant.

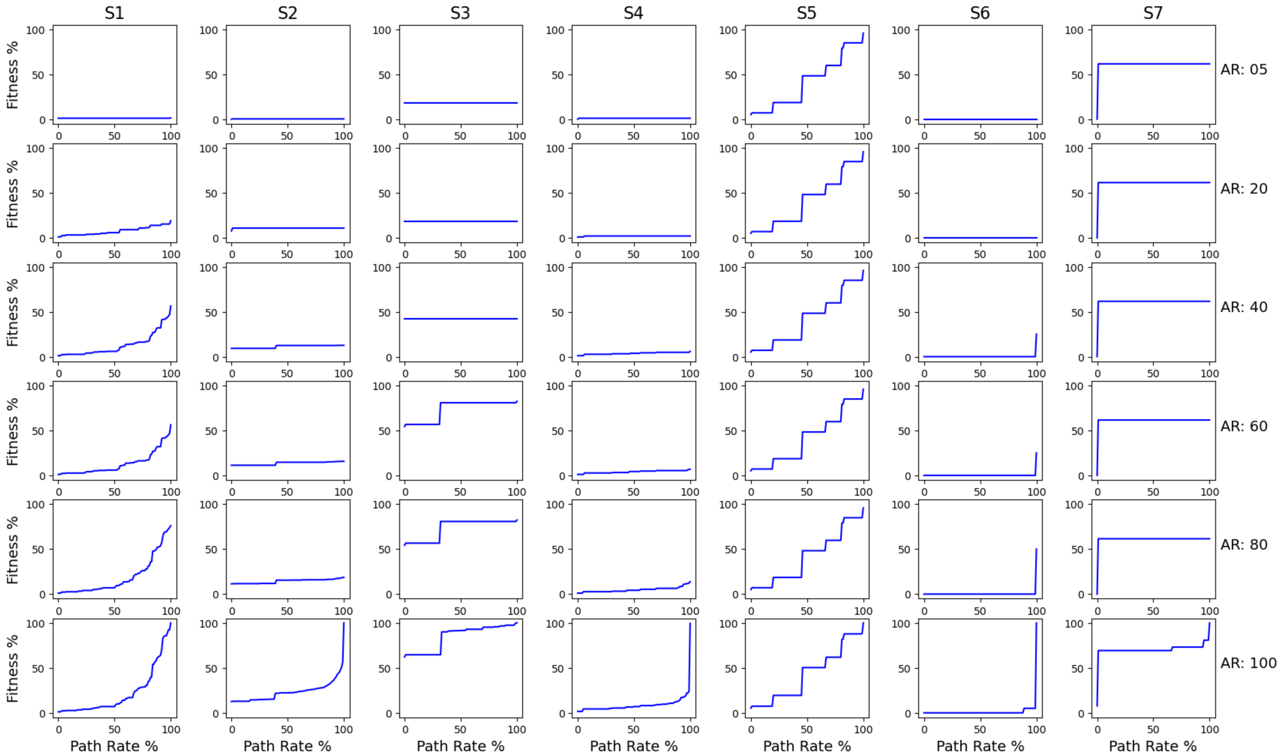


Figure 7. Fitness ratings of datasets 1 to 7 in R2.

A measure of the amount of behaviour only possible in the model and not in the event log is called Precision. Figures 8 and 9 show the behaviour of precision in R1 and R2, respectively.

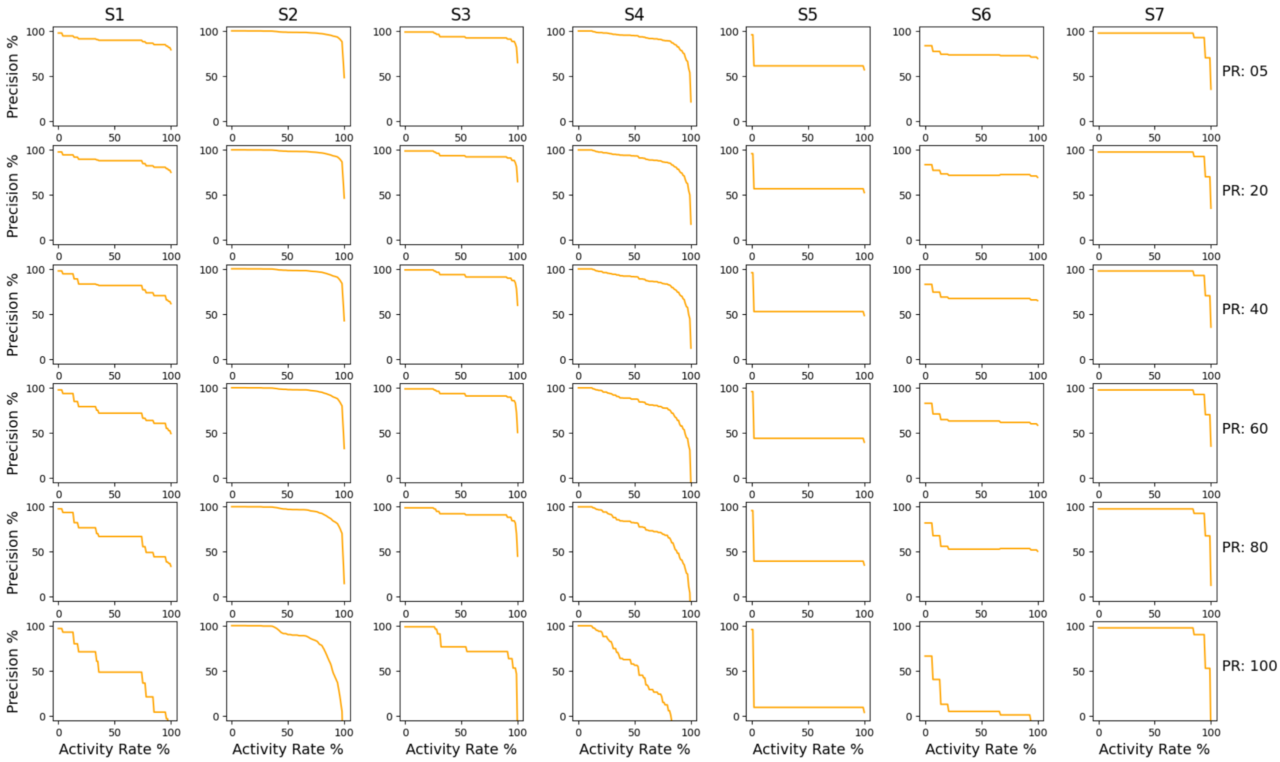


Figure 8. Precision ratings of datasets 1 to 7 in R1.

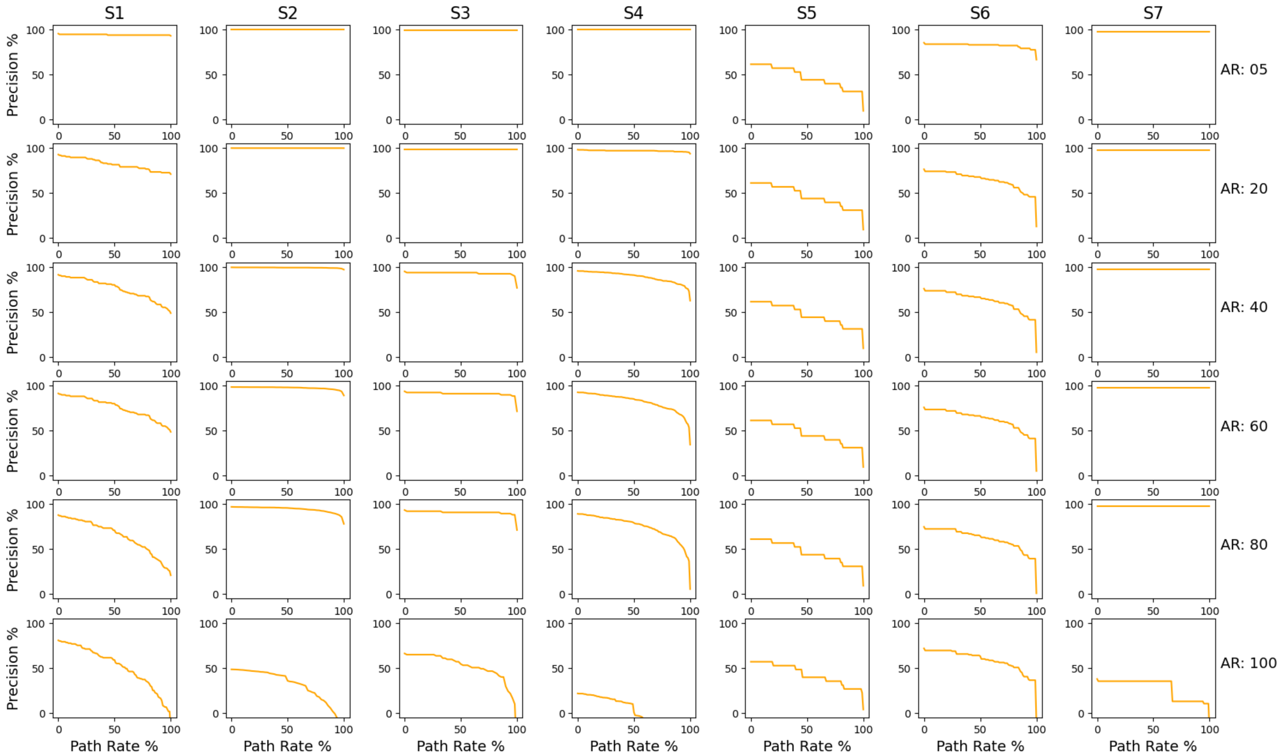


Figure 9. Precision ratings of datasets 1 to 7 in R2.

In general, the precision for all the datasets is decreasing in R1. In Figure 8, activity rate gradually impacts the precision of the model. As the path rate increases, the precision reduces. All the samples follow consistent behaviour of gradual decrease in precision with varying step sizes. The precision reaches the lowest value when both path rate and activity rate are at 100%. This is because, despite showing all available traces and events, there are still many possible combinations that were not shown in the log but are theoretically possible in the model. The slow change in precision is seen in path rates of smaller magnitude; for path rates with higher magnitude, the precision reduces faster since the number of possible combinations increases exponentially. Similar to R1, the precision in R2, shown in Figure 9, reduces as it approaches 100% of activity rate and path rate. The rate of precision reduction in R2 is lesser than in R1, as the number of activities does not have a major impact on creating paths for new combinations.

Generalisation is a measure of the level of abstraction that takes place in a process model. A model with uniform usage across all parts has good generalisation, while a model with infrequent usage across some parts has bad generalisation. Figures 10 and 11 show the behaviour of generalisation in R1 and R2.

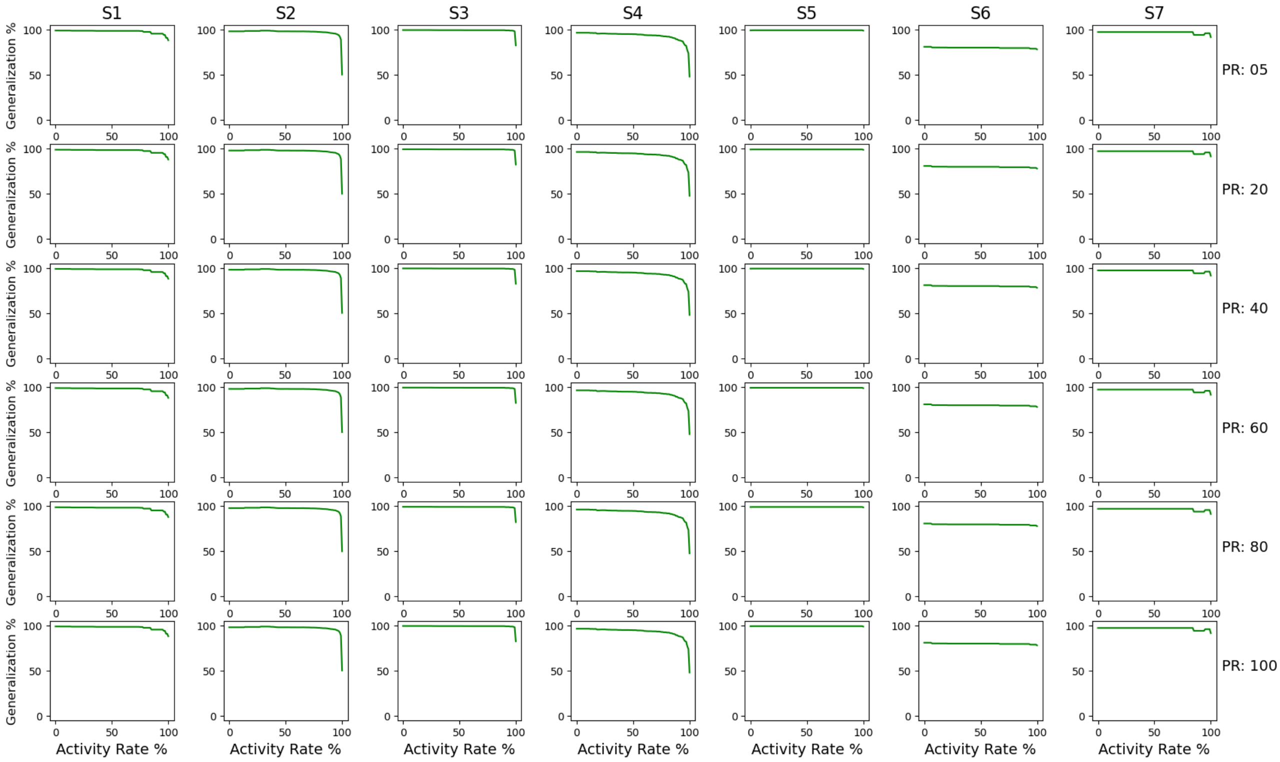


Figure 10. Generalisation ratings of datasets 1 to 7 in R1.

The behaviour of generalisation across all datasets is similar in R1. The generalisation rating is constant for majority of the time, which indicates all parts of the model are being used consistently. When reaching 100% activity rate, there is a sudden drop observed in S2 and S3, while S1 and S4 have a gradual decrease in their generalisation. S5, S6 and S7 are mostly constant throughout, with minor fluctuations nearing 100% activity rate. The variation in generalisation indicates the addition of an event into the model. In S1 and S4, the events being added after 50% activity rate are less frequently used; hence, all parts of the model are not equally used, thereby reducing the overall generalisation of the model. In S2 and S3, the huge dip in generalisation is due to the appearance of sparsely occurring events added at 100% activity rate. These events cause the process model to be less generic, resulting in low overall generalisation rating. In S5 and S6, the frequency of events have very similar thresholds. This causes most of the events to be shown within very small intervals of the activity rate. Hence, the number of activities remains constant throughout. This also means that all the activities are equally utilised in the event log itself, thus making the process model generic.

The activity rate controls the number of activities shown in the process model; therefore, an increase in activities impacts the generalisation based on the frequency of occurrence of the events. When frequent events are shown, the generalisation is high, since the events are used consistently throughout. When infrequent events are added, the generalisation is low, since certain parts of the model are not always used. This can be observed in Figures A7 and A8, where the number of infrequent activities correlates to the reduction in generalisation. The red circles indicate the new events that were added at each iteration of activity rate in R1. In our adaptation of this metric, the path rate does not have an impact on the generalisation in R1, since we account only for the activities and not the edges. By adding edges for the calculation, the results might not be conclusive, as the concept of generalisation is designed to measure the events in the model that is most often used. The addition and removal of paths will not have a direct effect directly the threshold of activities is maximum.

The behaviour of generalisation in R2 is straightforward, as observable in Figure 11.

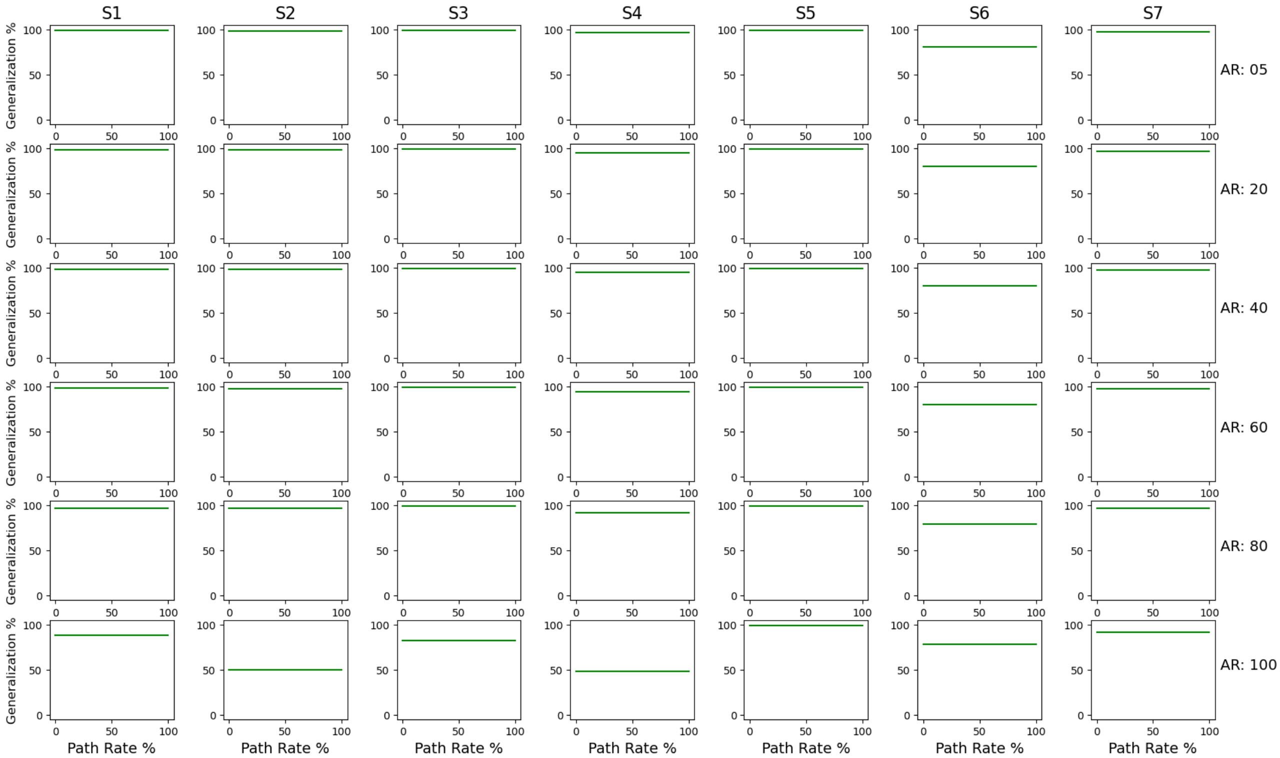


Figure 11. Generalisation ratings of datasets 1 to 7 in R2.

The generalisation stays constant throughout the increase in path rate because the number of activities is fixed in R2. E.g., at activity rate = 5, only 5% of the most frequent activities are shown. Thereby, the model will have a fixed number of events throughout the variation in path rate in R2. In S1, at activity rates 5 to 80, the generalisation is almost maximum. This is due to the total number of activities being constant throughout the analysis. All the events are used in the model equally throughout R2. At AR = 100, the generalisation decreases, due to the addition of all paths present in the event log. These paths are not utilised as much as the major ones; hence, the model is less generic than before. Figures A9 and A10 showcases this behaviour. Similarly, in datasets 2–7, each stage of activity holds the total amount of events constant. Therefore, the generalisation in R2 has no impact on the change in path rate and only changes when the activity rate is varied by its stages.

Simplicity is a measure to understand how easily the model can be perceived by a human subject. Figures 12 and 13 show the behaviour of simplicity in R1 and R2. The behaviour of simplicity is equated to the complexity or difficulty in understanding the model. The greater amount of events present in the model, the more difficult it is to understand for a human subject. In this adaptation of Simplicity, the number of interconnections was not involved in calculating the metric. Hence, in R2, despite the change in path rate producing a change in the model, the simplicity remains at the same level, (i.e., due to the mathematical formulation not accounting for the paths).

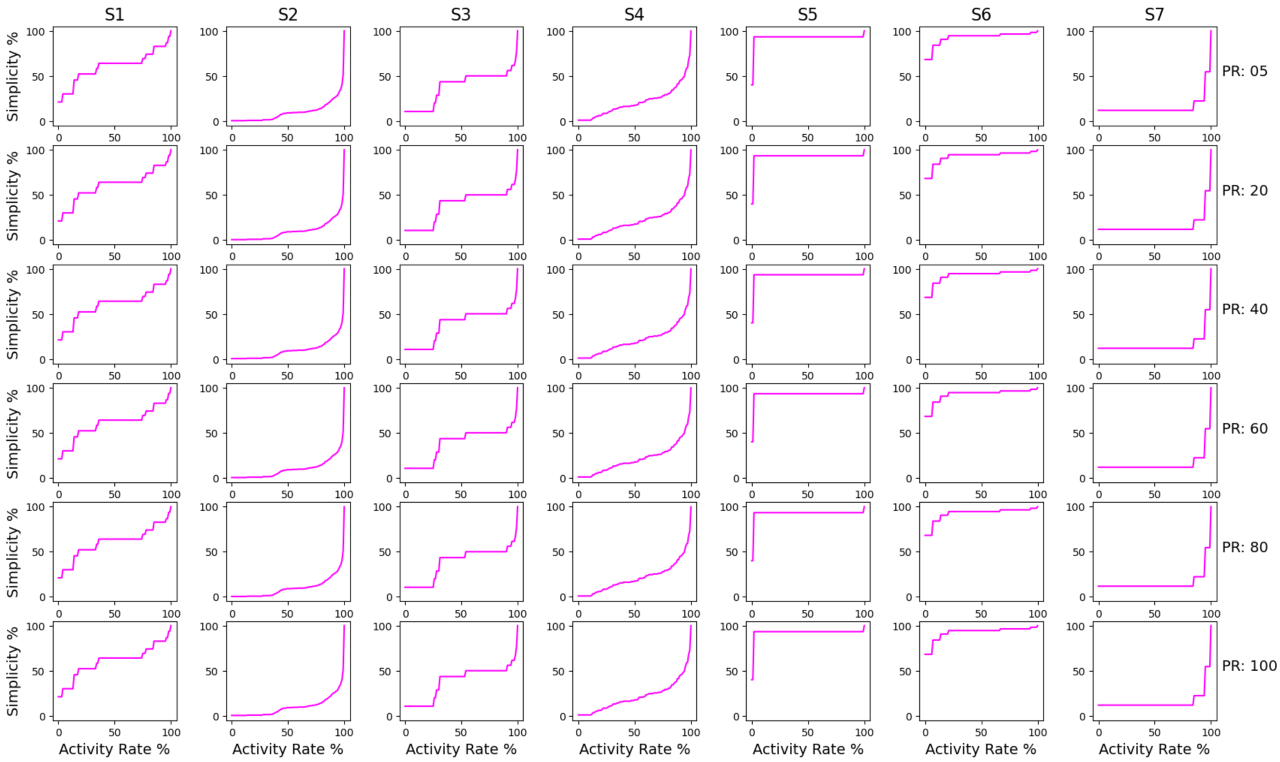


Figure 12. Simplicity ratings of datasets 1 to 7 in R1.

In R1, the behaviours of the datasets have an increasing trend with gradual steps and constant behaviour. Samples S1, S2, S3, S4, S6 and S7 have a step-wise increasing behaviour for simplicity. These steps are instances when new events are added to the process model. Thus, increasing the number of activities and edges impacts the complexity of the process model. The Figures A11 and A12 show an example of S7 in R1. The increase in activity rates at path rate 100 correlates to the increase in complexity, as in Figure 12 (i.e., the rapid change in complexity is seen only at the end process models, visible in Figure A12). S2 and S3 have a smoother curve, which is due to the presence of many activities with varied thresholds.

S5, as seen in generalisation, has many activities with similar frequency counts, thereby passing the threshold at the same time. This results in many of the activities being shown most of the time and no addition of new events except in the extremes. Figure A13 shows an example of when the model stays constant throughout, from activity rate 5 to 95% at path rate 100%, after which there is a change in the model and it continues to remain constant.

Simplicity in R2 is similar to the behaviour of generalisation in R2. Its behaviour is shown in Figure 13. In all the datasets, the number of activities remains constant throughout the cycle of path rates from 0 to 100. With the number of activities remaining constant throughout, the number of paths is the only feature that can be increased. However, since the paths can only be drawn for existing activities on the process model and it not being included in the calculation, all the models have constant activities but with varying levels of interconnections. Hence, this version of the simplicity formulation produces a constant behaviour in R2, yet, if a formulation is created to measure the paths, it would take into account the detail being produced from the interconnections. The behaviour can be seen in Figure A13; dataset 5 is in R2 and the simplicity remains the same throughout.

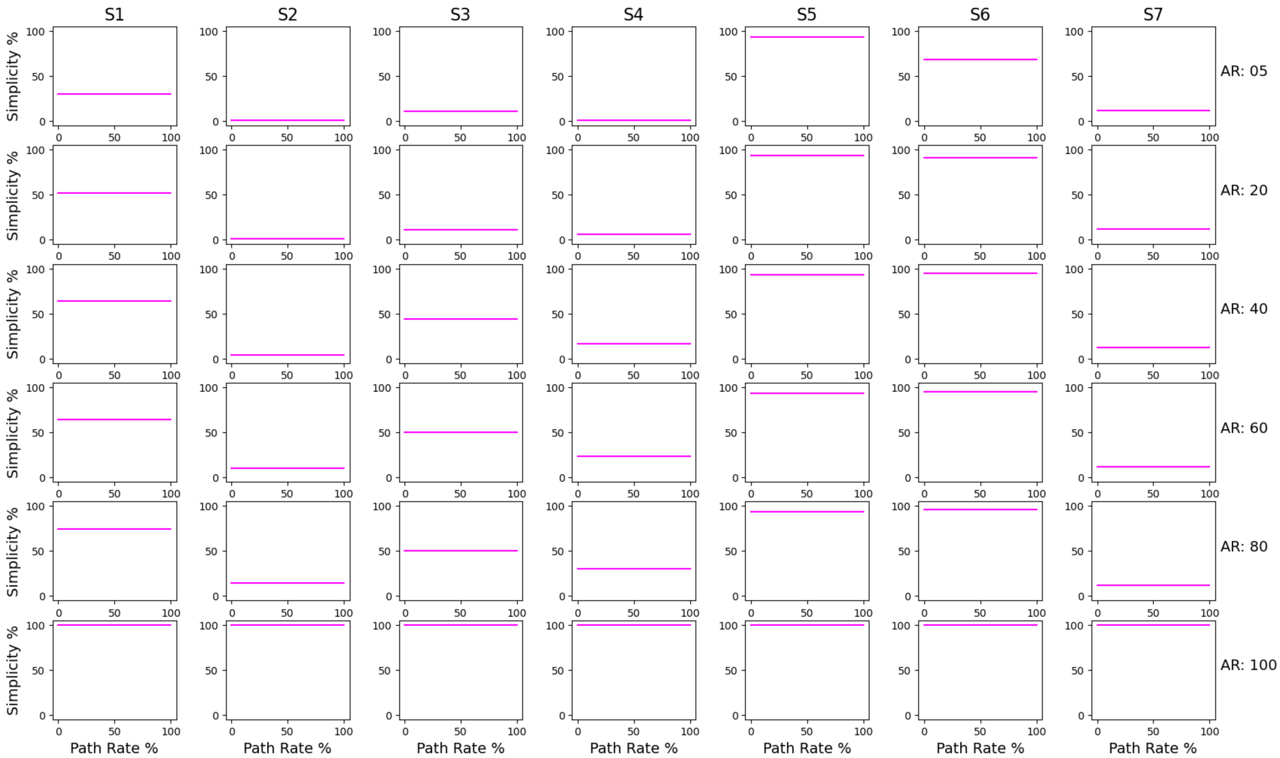


Figure 13. Simplicity ratings of datasets 1 to 7 in R2.

Further, we have included full-scale process graphs for select datasets in R1 and R2. Due to the limitation of size and detail, only four datasets were compact enough to be reproduced on sheets. Appendixes A14–A21 show full-scale process models for datasets 1, 3, 5 and 7, respectively. The process maps can be correlated with the behaviour of the QUAD metrics in Figures 6–13 in both R1 and R2. When observed horizontally (from left to right) the process maps are in R1 order and when observed vertically (from top to bottom) the process maps are in R2 order.

4. Discussion

Our work provided a ton of insight into metrics and the evaluation of process models. In this section, we have listed important issues that were encountered during this work. Addressing these issues will further improve the understanding of process models.

Metrics, preferences and their application. Every person is accustomed to their preferences. Complex process models are suitable for industry professionals, as they portray information in greater detail. However, beginners would find complex models very hard to understand. Using a constant model for both audiences results in differences in detail being experienced by either party. This causes a loss of information. To answer this issue, there are several ways we have considered. The first method would be the generation of personalised models depending on the intended user's preferences, background and subject knowledge. The possible backdrop of this method would be the aggressive collection of personal data. A more realistic approach involves creating a universal model that describes a process within a balance of complexity (i.e., a model that can be useful for both beginners and experts). Metrics and evaluation techniques can be used to determine the optimal range of information to be described via the model. The accuracy of this range can be further improved by incorporating secondary data from the process model (e.g., size of traces, avg. size, number of unique events, etc.).

Figure 14 describes the behaviour of QUAD metrics when an event log is tested in R1. This graph explains the performance of the model and, by analysis, we find the optimal range where models have a balance between all the metrics.

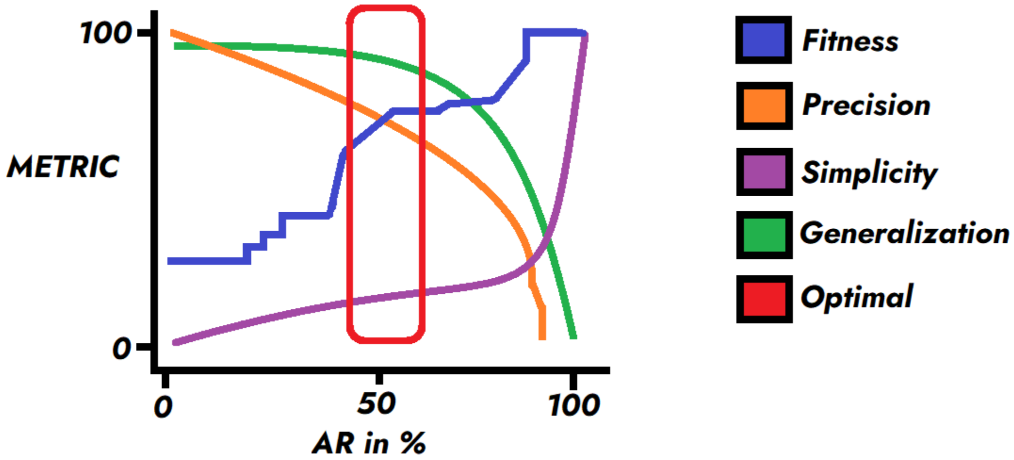


Figure 14. An optimal range for a process model in R1.

At activity rate 50%, it can be observed that the fitness is greater than 75%, indicating that 3/4 of the traces from the event log can be replayed by the model. The generalisation is good, signalling that most parts of the model are used equally. The precision at the same point has started to reduce but is greater than 75%, signalling that the model might have a small amount of behaviour unseen in the event log. The simplicity is low, indicating that the model is easy to interpret. From these observations, it can be concluded that the region near the 50% mark can be termed as the optimal range, since the models generated within this area during simulation have a good balance between all metrics. This result proves that the model is not aligned with any party (expert or novice). Hence, by using metrics, balanced models can be estimated and used for universal purposes. When required, individual models with weightings on specific features such as complexity, fitness, etc. can be explicitly created for usage (e.g., if all traces should be shown in the model, then the point at which fitness is 100% should be chosen).

Looping in precision. During our experiment, we encountered an issue while calculating precision. It is best explained with an example. Consider a process model with the following flow {AB, BC, CA, BA}. To calculate precision, we calculate the various possible paths between two points. To evaluate, we utilise the full trace of individual caseIDs to check for behaviour not seen in the event log. In the exemplar process model depicted in Figure 15, the model starts at A and ends at C. It has one intermediate event B.

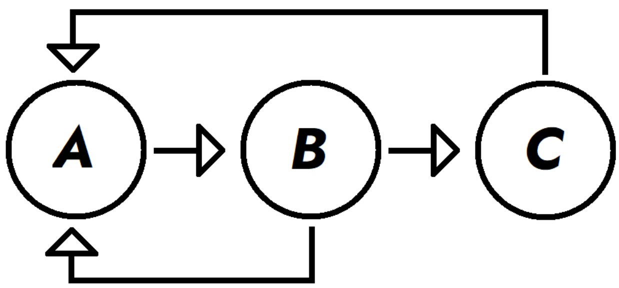


Figure 15. Process model to demonstrate the looping threshold.

We can notice there exists a looping construct BA and CA. The presence of loops gives rise to uncertainty, i.e., *How many times should an occurrence of a loop be included for evaluating precision?* For example, a trace in this event log could consist of the sequence ABABABABABC as one of its entries. If we set our looping threshold to three (i.e., any activity can repeat only three times), the precision of this example would be negative. This is because of the existence of duplicate events that are not recorded, giving an impression of too much existing behaviour. This can also be seen in our experimentation of precision in R1. Some samples cross the 0% mark and start generating negative precision values. This is due to the same issue [97–99].

Specific changes occurring due to variations in the event log during process map creation. This is an example of the inclusion and exclusion of events causing changes in the metrics recorded. In our experiment, we faced issues with our datasets when using "START" and "STOP" events within our models (i.e., the trace begins with an event "START" and ends with "STOP"). Some of our datasets did not explicitly have this event within them, while a few did. We decided to remove the START and STOP events and then evaluate our metrics. Before taking this decision, we analysed the possible issues that may arise by including and excluding these events.

By excluding START and STOP

- The traces are uniform for universal analysis;
- Traces begin and end with activities from the log, resulting in directly analyzable pure information;
- By excluding START and STOP, instances of measurement are scaled exactly to their logs and not the collective size.

By including START and STOP

- In datasets without the keywords, traces have to be explicitly extracted and modified, thereby increasing pre-processing time;
- When there is more than one event that leads to a STOP, there is an imbalance in precision and generalisation, since they would be additional events that individually (tracewise) lead to the end but collectively result in a mismatch [100–103].

Why choose understandability and complexity? From our study, we find that the main motive and objective of process modelling is to present event log data to humans in a simple form. The quality of complexity covers many terminologies that give the idea of *"not difficult"*. Understandability gives the idea of "*easy to understand*". These two qualities appreciate each other and provide a certain platform to clearly state whether a model is simple or complex without forgetting about metric constructs [104].

Emphasis on using pre-processed data for analysis of events: This is a crucial issue. As mentioned earlier in Section 2, the basic three are vital for process mining and modelling. However, the task data should undergo careful pre-processing, as this can cause issues when evaluating metrics. In our experiment, we faced an issue when there was the usage of English and Cyrillic alphabets together. Our pre-processing was designed for filtering and cleaning English alphabets. However, we overlooked the Cyrillic alphabet which led to more events being present than in the event log, resulting in process models with an increased number of activities to be generated. The evaluation of models with proper per-processing was found to have an equal number of activities as in the event log, thus avoiding irregular metric ratings.

Limitations of the work: By performing this study and experiment, we found certain limitations in the technical aspect. In our work, we have focused on three human-centric qualities (understandability, interpretability and complexity). However, many more qualities can be explored and used to classify metrics (e.g., modifiability, expressiveness, compatibility, etc.). Classifying metrics based on such qualities requires an increased depth of research and narrow exploration strategies.

ProFIT process mining toolkit currently works only using the fuzzy heuristic mining algorithm and visualises data using the direct flows graph (DFG). In comparison to other mining algorithms and visualisations, there are limitations with using this technique (e.g., the Petri-nets can show more event data that is hidden in DFGs and other mining algorithms may project the data in a different perspective). Hence, there is an opportunity to explore the various mining and visualisation techniques; however, our experiment focuses only on the behaviour of the fuzzy heuristic miner and the DFG visualisation.

In our adaptation of the QUAD metrics, we noticed that Simplicity and Generalisation produce constant graphs in R2. At first, we assumed this was an error. However, the reason for this behaviour is due to the non-inclusion of interconnections (edges) from the model into the calculation of the metric (i.e., the paths connecting the model are not used in the calculation of the event log). From our experimental results, we plan to improve these metrics and produce a version that utilises all elements of the process model for evaluation.

The source code created in Python3 for cleaning the data was specifically crafted only for the selected event logs. To use new event logs with variable characters present, a universal filtering system should be created for each event log.

In our work, we have focused on using medical datasets as the prime domain to showcase the usage of QUAD metrics and their behaviour. Nevertheless, the metrics are expandable to all domains with adaptations. By using the QUAD metrics, major characteristics of medical procedures can be identified and optimised. For instance, the generalisation metric can verify the frequency of elements in the medical procedure and eradicate operations that may increase the duration of the process. Fitness can be used as a scale to visualise the whole process and analyse the complete medical procedure without the need to extract metadata of the operation. Similarly, other metrics can be used to identify individual characteristics of medical procedures that allow for the optimisation and restructuring of processes. By applying these aspects, medical procedures can be improved, therefore, benefiting the medical community and, ultimately, the end user (i.e., the patient).

In our experiment, we observed instances where the process models had bottlenecks and peculiar operations that impacted the overall interpretability of the model. For example, in dataset 5, the process models present in Figures A18 and A19 show that the events *"Triage"* and *"Register"* occur in almost all the process models but, with increased activity rates, a third event *"Check"* tends to bottleneck all events taking place after it. This causes the medical process to be dependent on the *"Check"* event at every iteration and may also increase the duration of the medical procedure. An ideal solution would be to split this event into smaller sub-events that are relevant to the medical operations being performed, thereby improving the efficiency of all processes after. This change can also be viewed via the metrics where the complexity of the model will slightly increase at the expense of better interpretability, generalisation and fitness.

5. Conclusions and Future Work

Achieving a balance between complexity and interpretability is vital to fulfilling the goal of process mining. Process models can be understood by the end user only if they are fit for purpose. Hence, metrics and evaluation techniques mentioned in this work provide researchers with a guide to enter and dive into assessment techniques for process modelling. The taxonomy of metrics collectively cumulates all the viable measures proposed by fellow authors during the time frame of 1997 to 2023 and classifies them based on three human-centric process modelling qualities. We present a once-and-for-all solution to the vast amount of metrics in the community using the taxonomy. Researchers can use our work to find metrics and evaluation techniques instead of studying many articles to find the right fit. Together with the broad spectrum review, the taxonomy follows through by summarising the history of how to evaluate a process model and provides a steady learning curve for beginners in the domain. We have chosen the four most widely used metrics from the tracked literature and termed them the QUAD metrics. These metrics utilise universal attributes of a process model for evaluation and can be used to assess the vital properties of the model. The performance and behaviour of these metrics under various control conditions have been showcased using real-world healthcare event logs. The conclusions drawn from the experiment will aid researchers in the analysis and selection of metrics for future applications. Our implementations of the QUAD metrics are adaptable to datasets from all domains with reconfiguration (i.e., they are universal). Modellers can obtain significant insight into the interpretability of their models by evaluating their process graphs with these metrics. Therefore, we propose to use the QUAD metrics as a universal standard of evaluation for process models. The insight from these metrics can be used to fine-tune and craft visualisations specific to certain qualities or audiences. We believe that our experiment provides crucial information that will aid future researchers to progress in the evaluation phase of modelling.

In future studies, we aim to create open-source access to our implementation of the QUAD metrics and introduce add-ons for universally used process mining software such as PM4PY. We consider the insights from our study to be generalised input for the community and scalable to all domains. We aim to resolve the limitations of this study by including more process modelling qualities for metric analysis and implementing re-conceptualised versions of simplicity and generalisation metrics to account for both activity and path rates. We plan to create a universal filtering script that is applicable to all domain data types within process modelling.

Author Contributions: Conceptualisation, S.V.K.; methodology, A.T.S.I. and S.V.K.; software, A.T.S.I.; validation, S.V.K.; formal analysis, A.T.S.I. and S.V.K.; investigation, A.T.S.I.; resources, S.V.K.; data curation, A.T.S.I.; writing—original draft preparation, A.T.S.I.; writing—review and editing, A.T.S.I. and S.V.K.; visualisation, A.T.S.I.; supervision, S.V.K.; project administration, S.V.K.; funding acquisition, S.V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Higher Education of Russian Federation, goszadanie no. 2019-1339.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the study are publicly available via links provided in the corresponding sections.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Metric Methodology Flowcharts

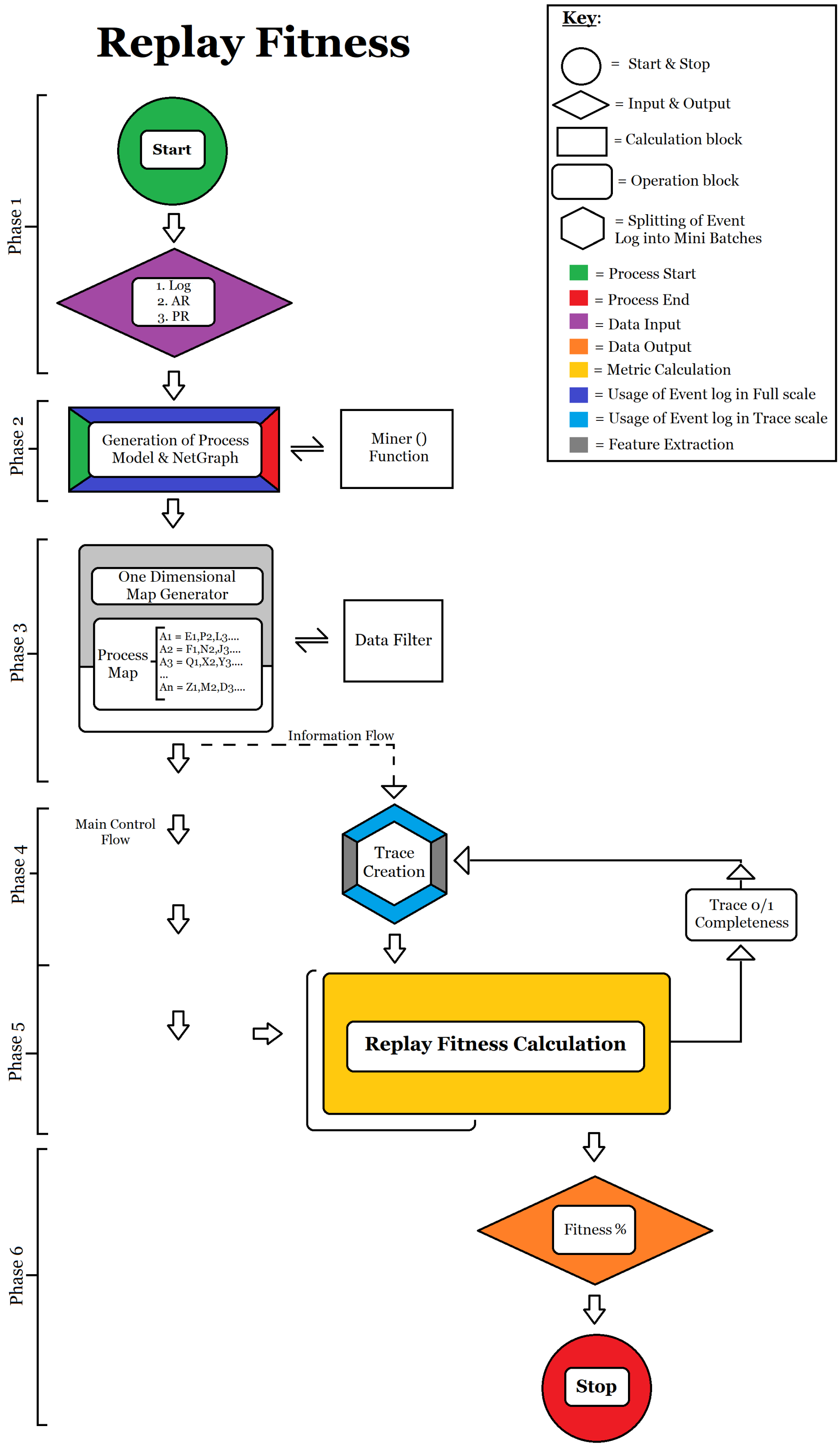


Figure A1. Workflow for Fitness.

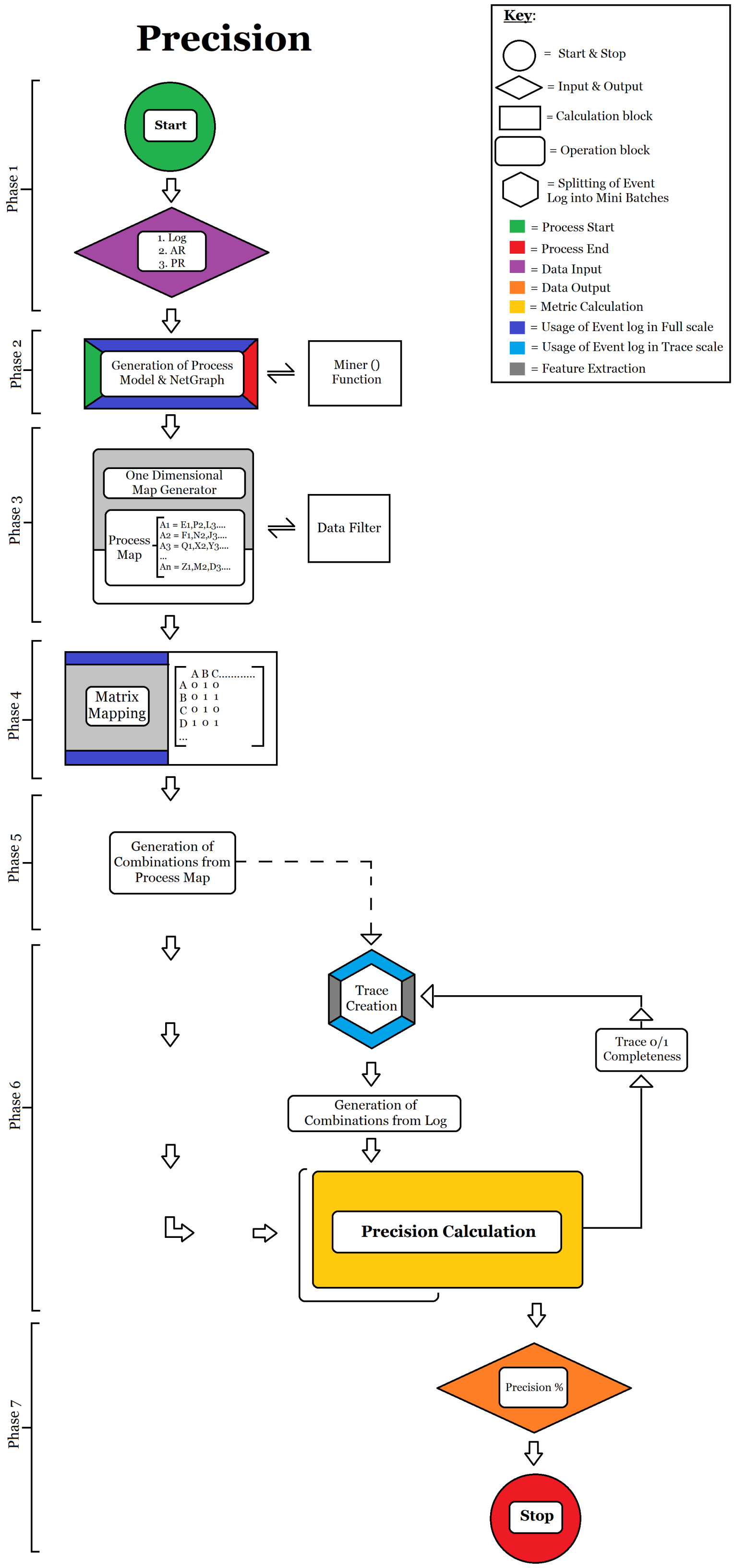


Figure A2. Workflow for Precision.

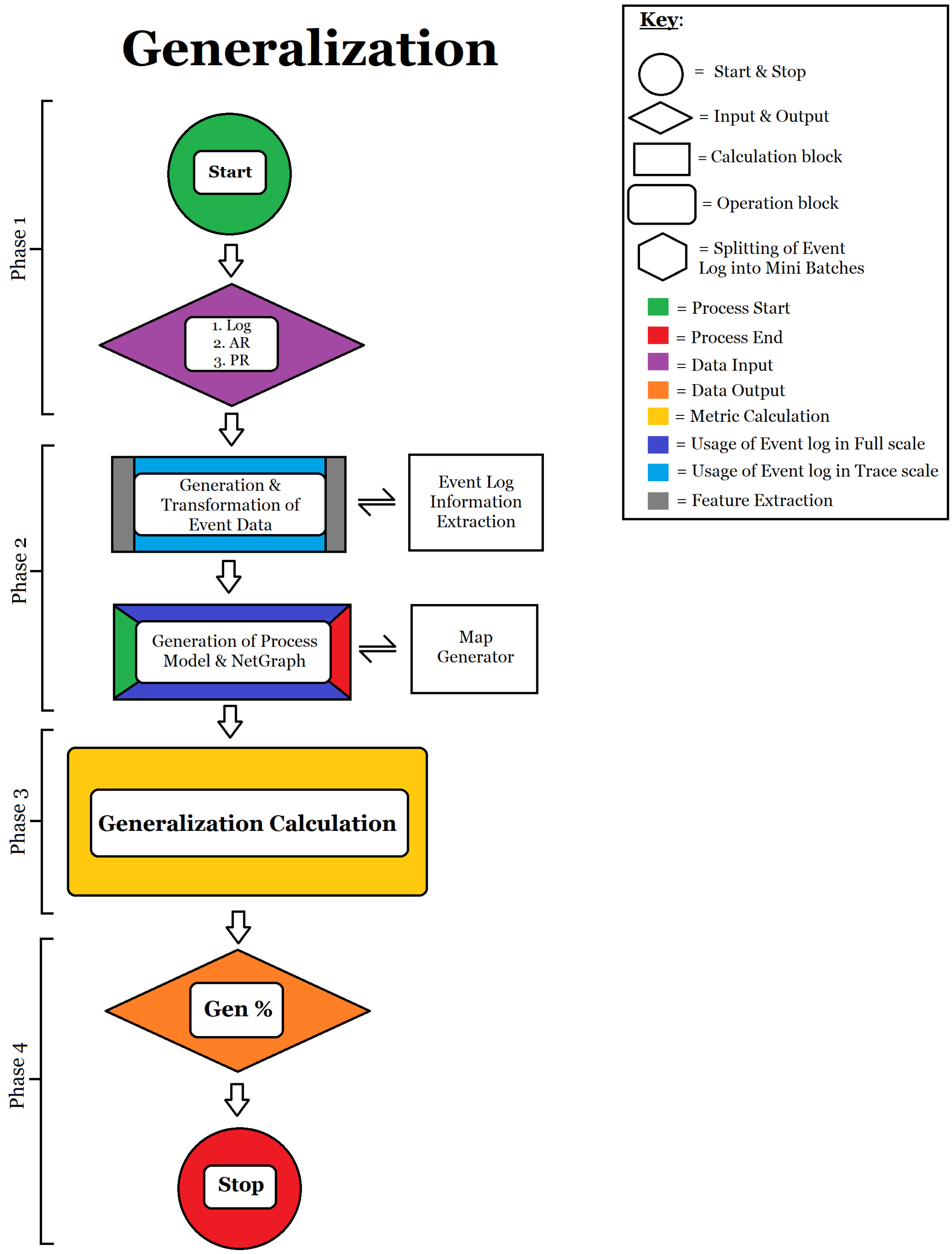


Figure A3. Workflow for Generalisation.

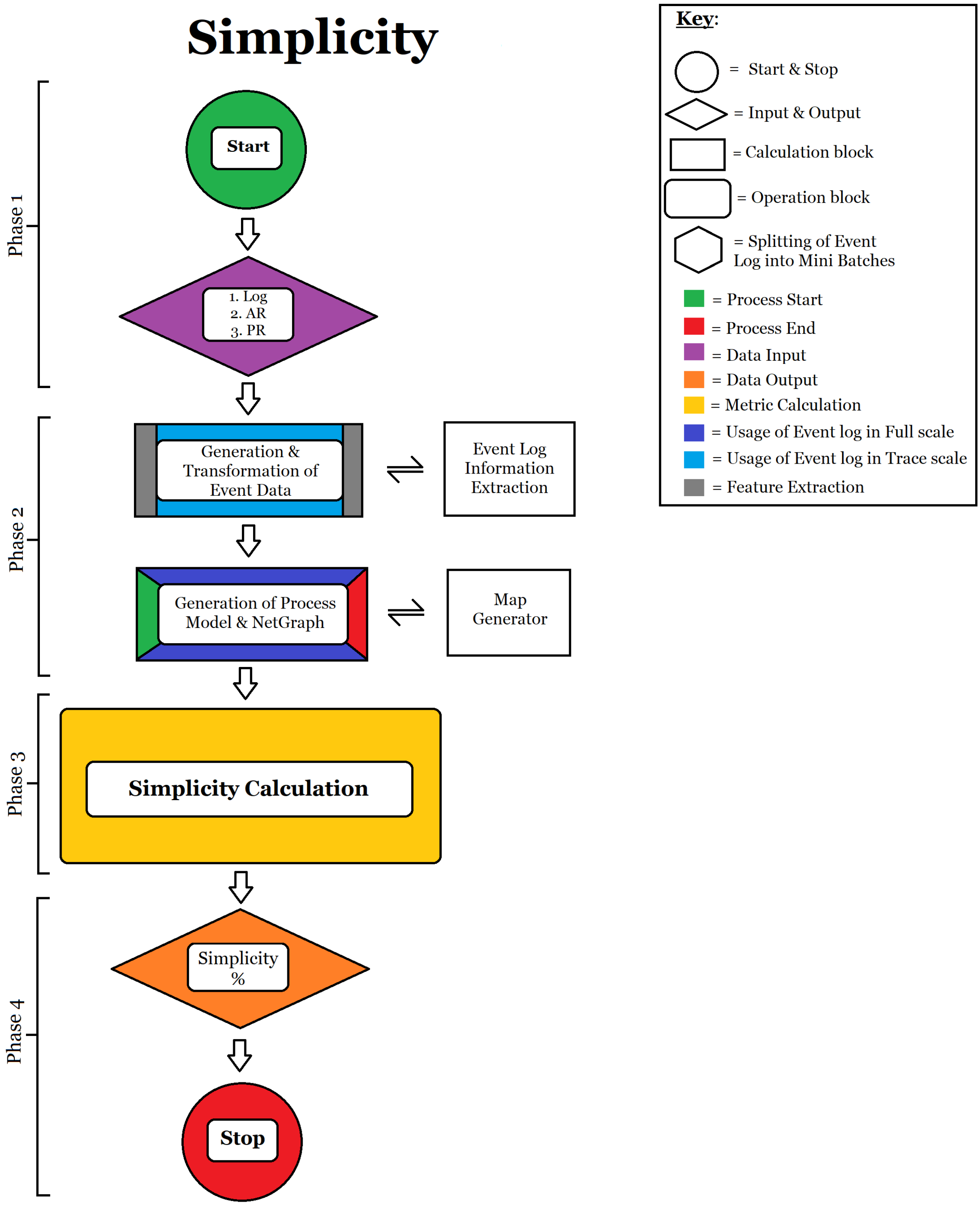


Figure A4. Workflow for Simplicity.

Appendix B. Process Model Graphs

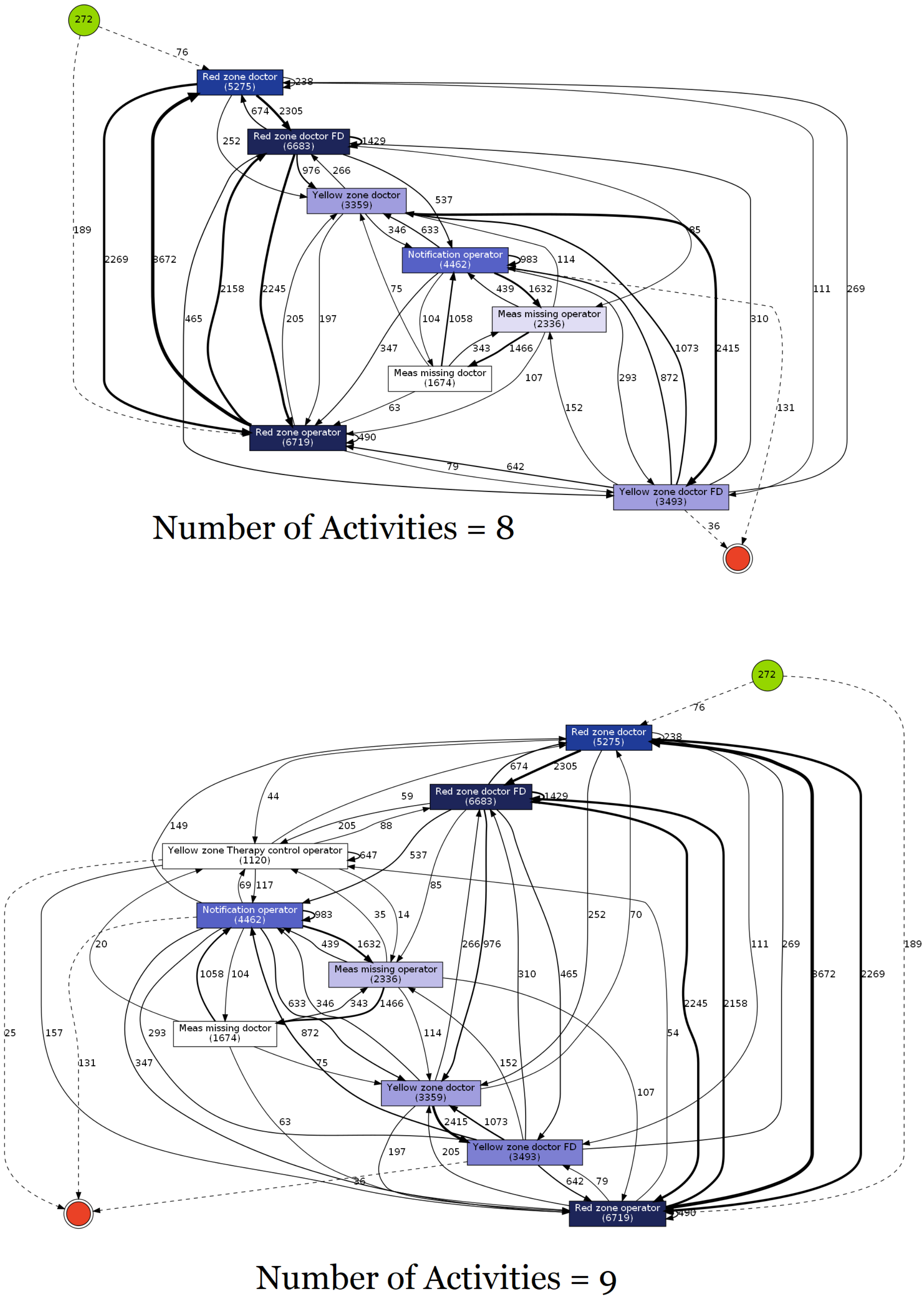


Figure A5. P1: Impact of adding activities on fitness—sample 1.

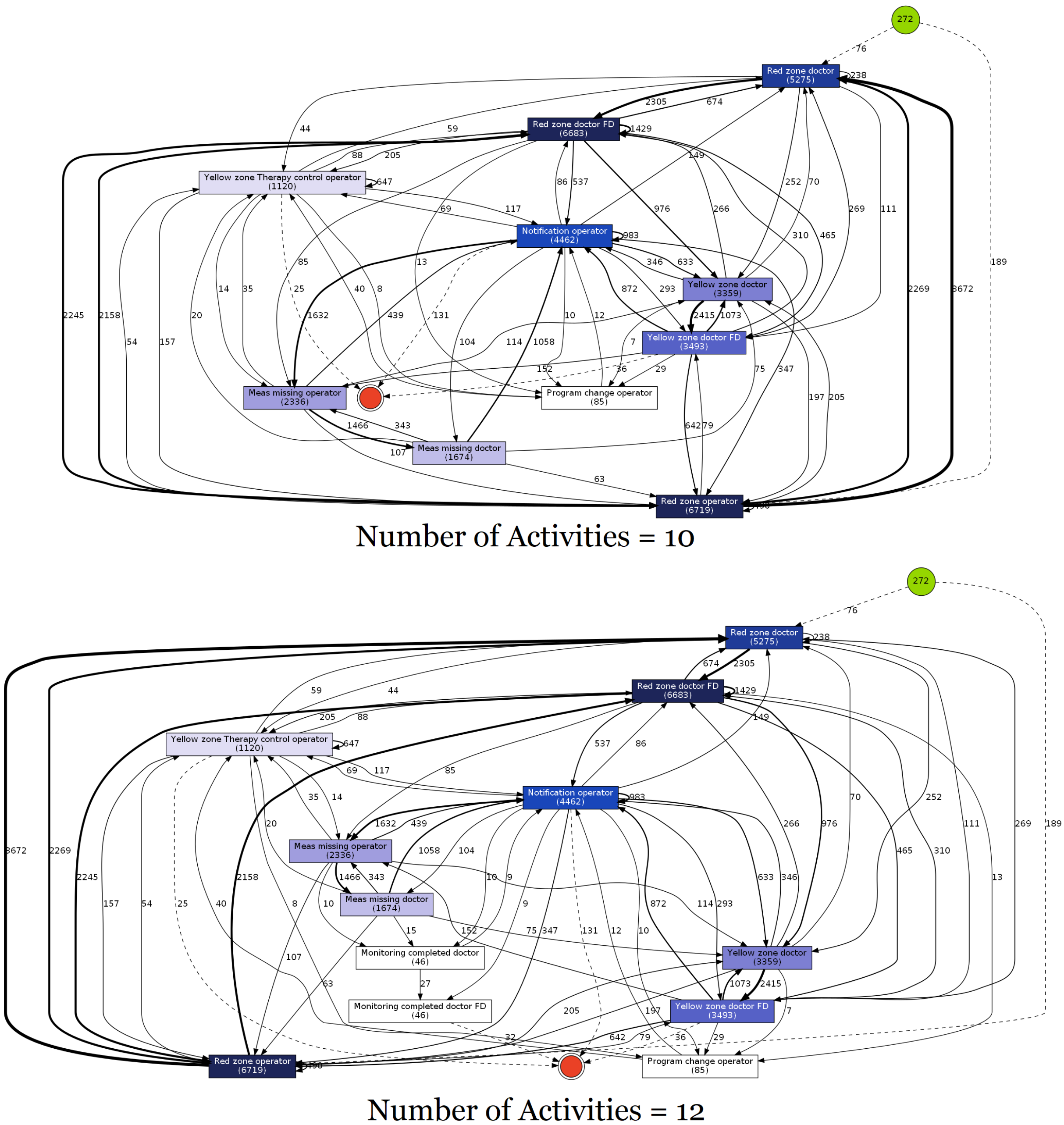


Figure A6. P2: Impact of adding activities on fitness—sample 1.

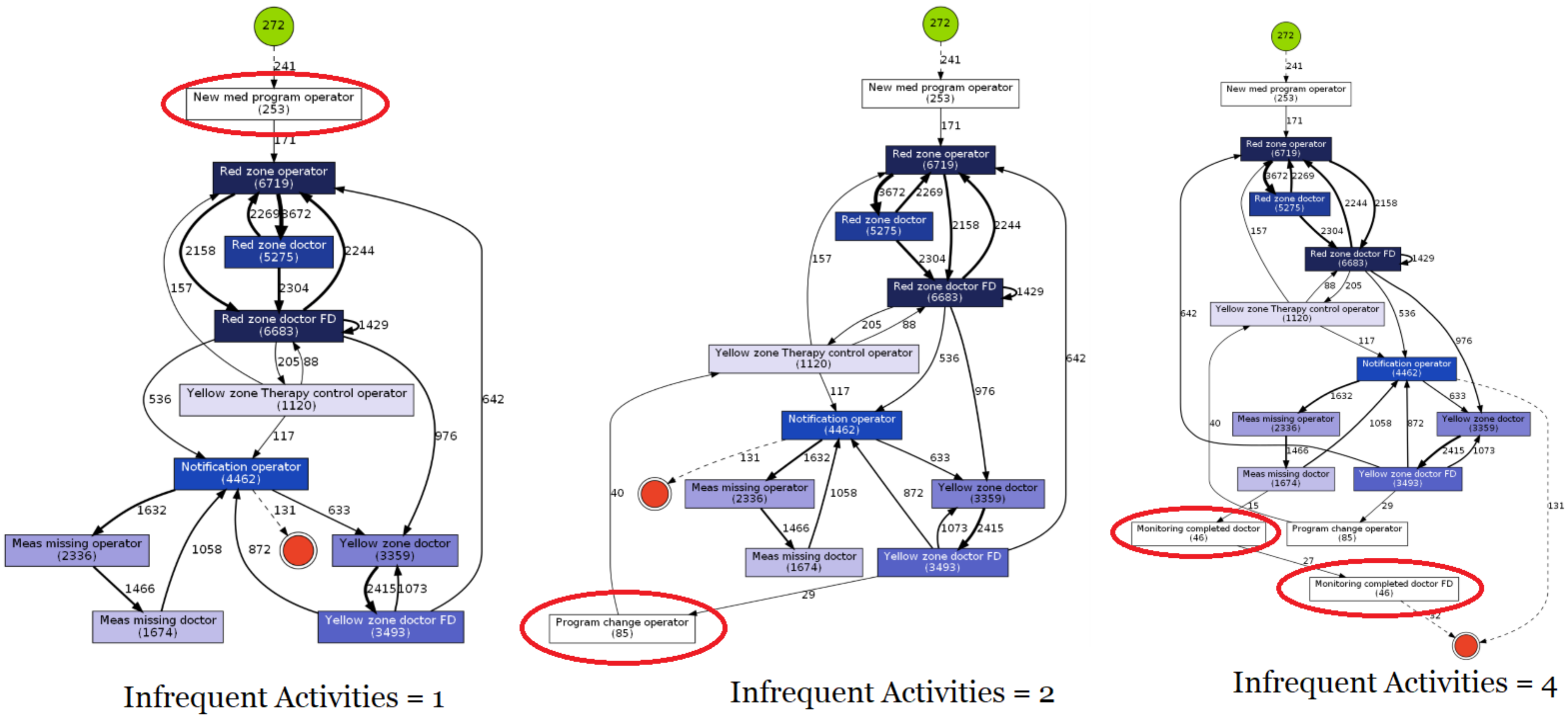


Figure A7. P1: Impact on generalisation due to addition of activities in R1—sample 1.

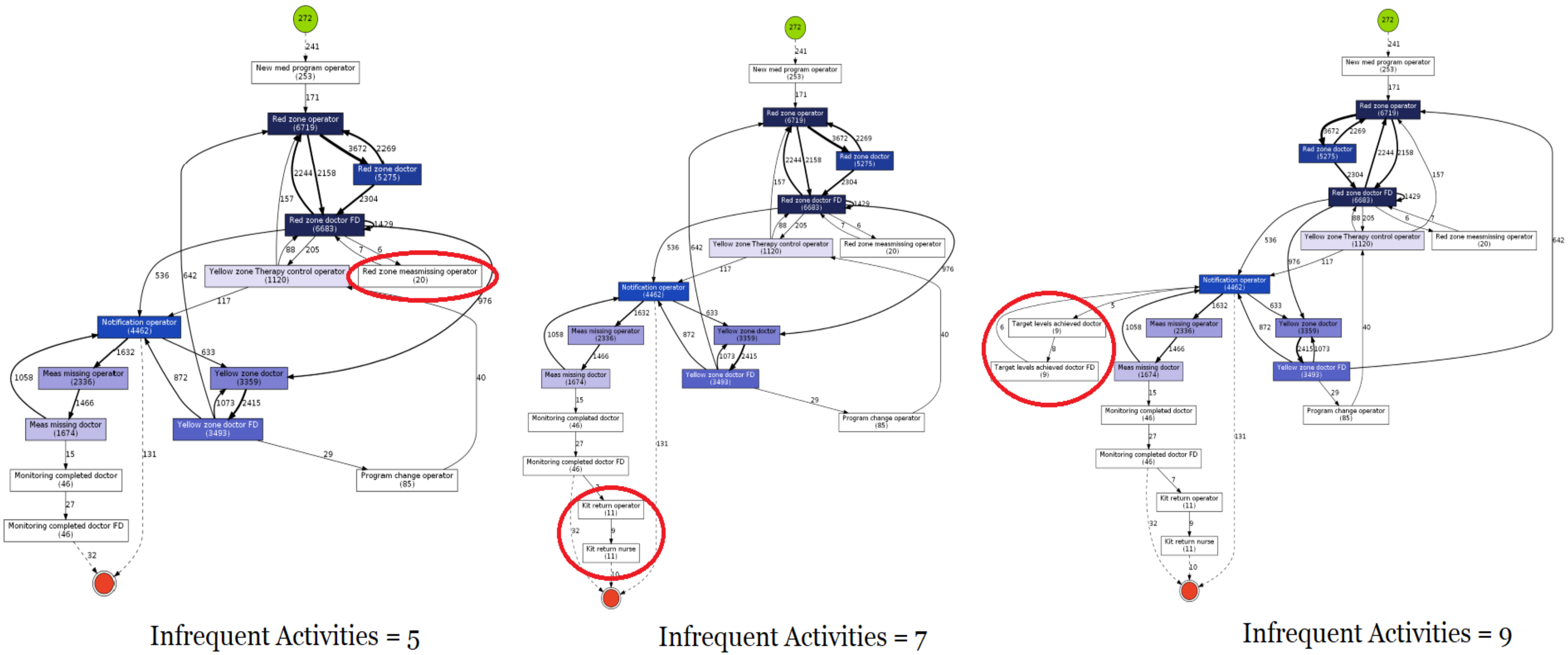


Figure A8. P2: Impact on generalisation due to addition of activities in R1—sample 1.

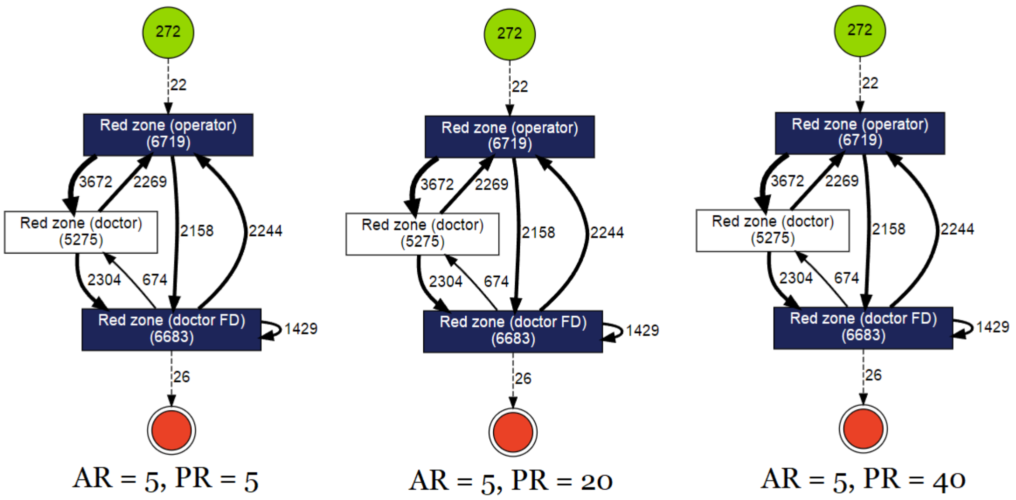


Figure A9. Impact on generalisation of the process model when in R2—sample 1–part 1.

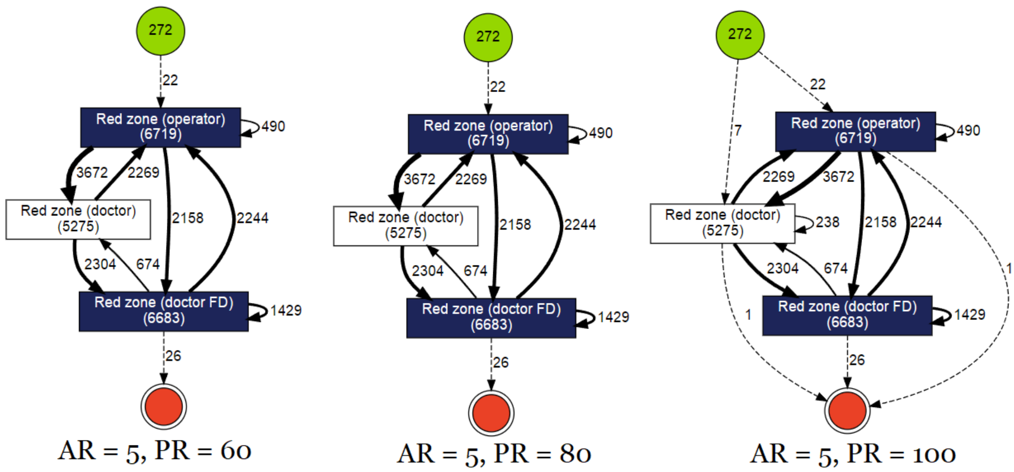


Figure A10. Impact on generalisation of the process model when in R2—sample 1–part 2.

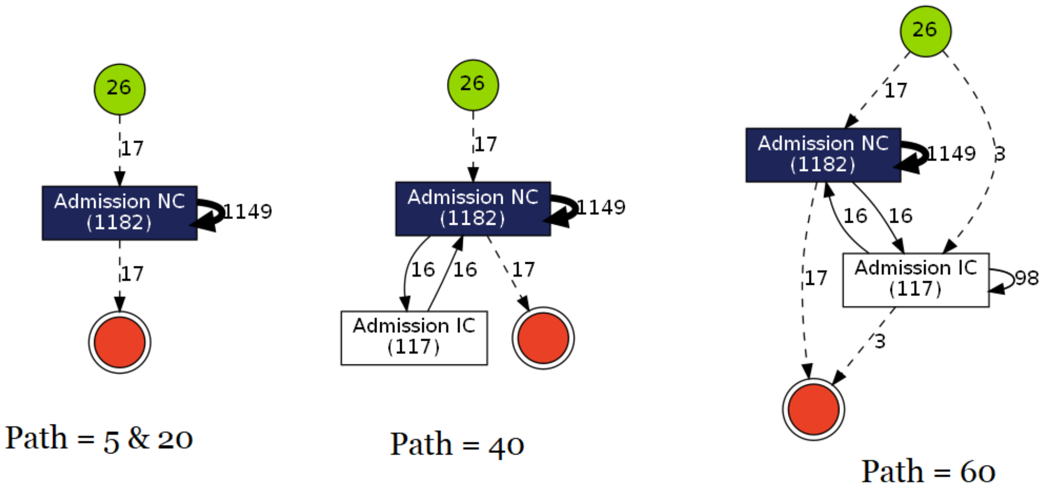
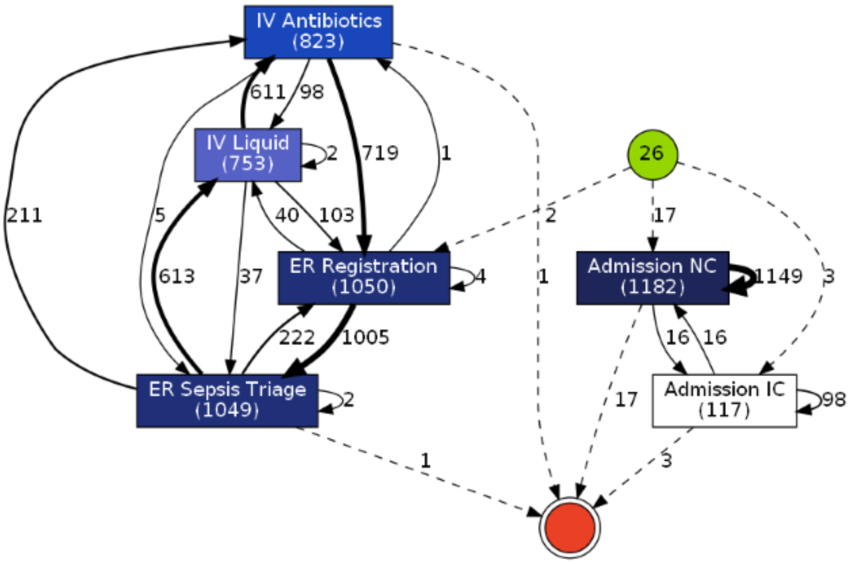
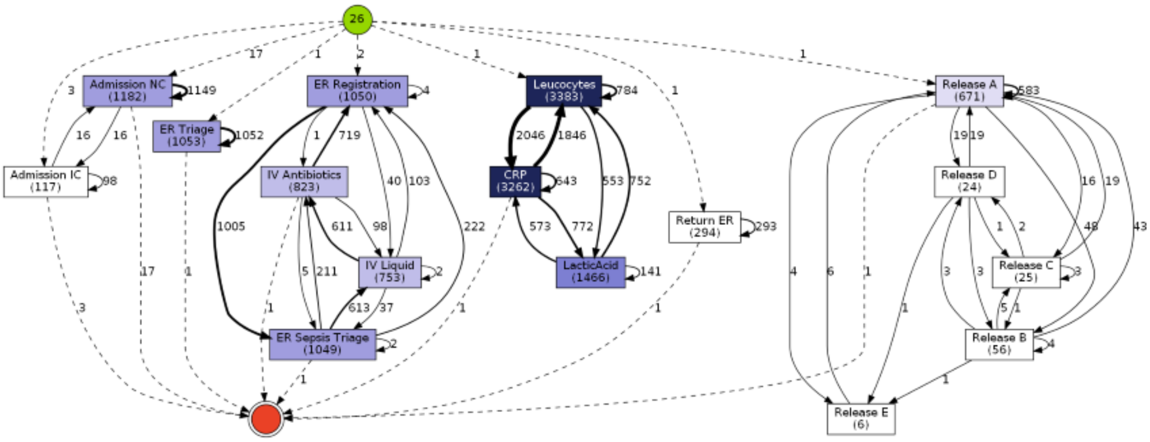


Figure A11. P1: Impact on simplicity in R1—sample 7.

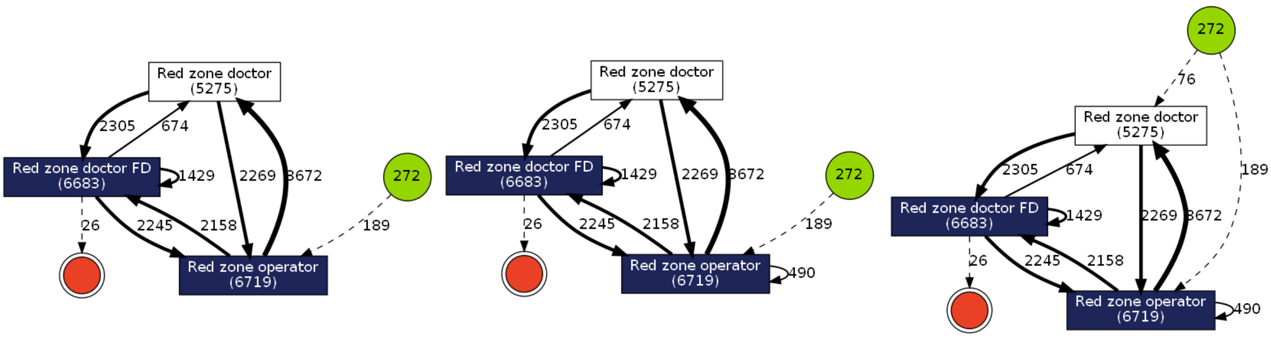


Path Rate = 80



Path Rate = 100

Figure A12. P2: Impact on simplicity in R1—sample 7.



AR = 5
PR = 5

AR = 5
PR = 50

AR = 5
PR = 100

Figure A13. Impact on simplicity in R2—sample 5.

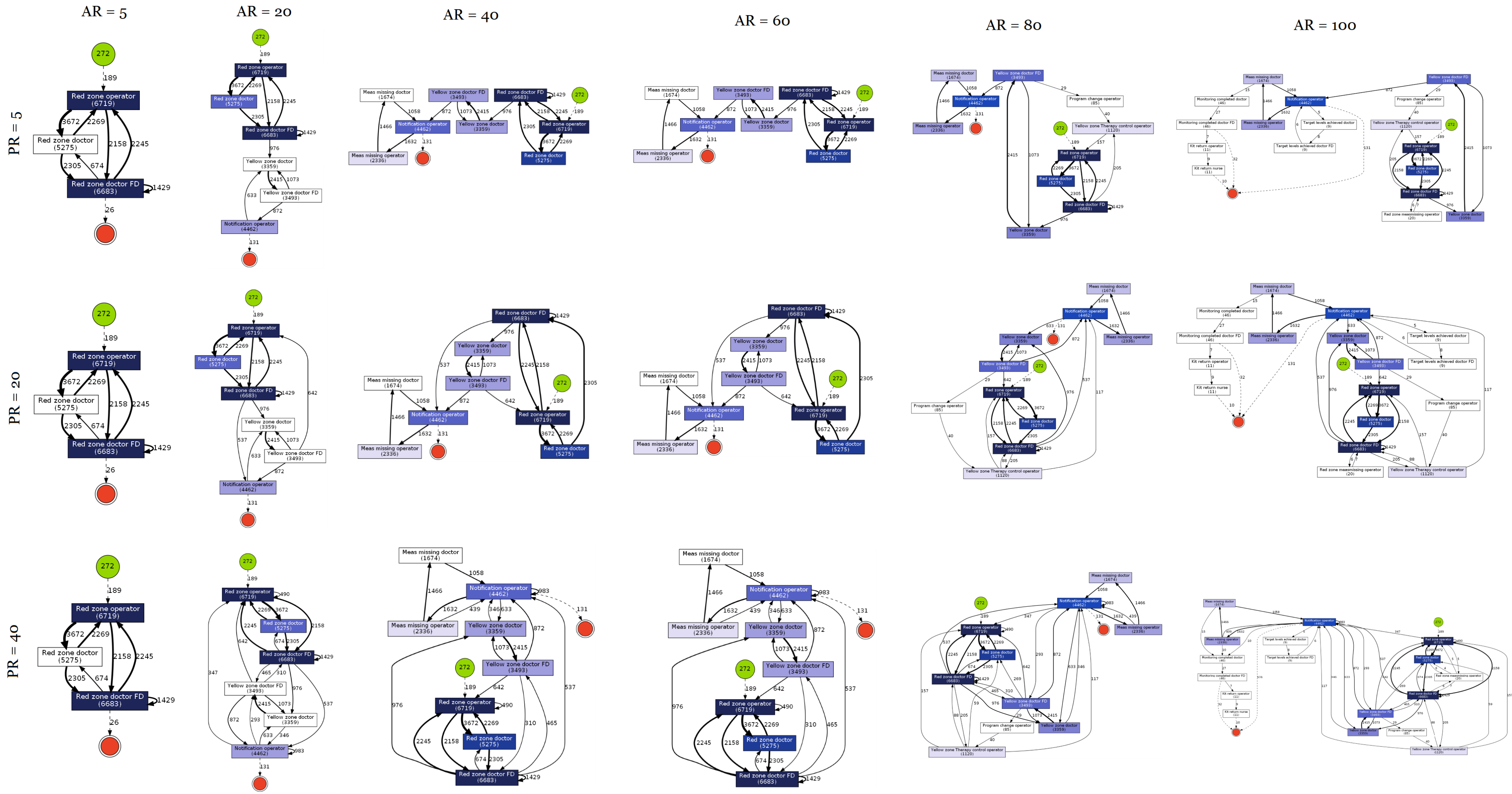


Figure A14. Process model graphs of dataset 1 in R1 and R2.

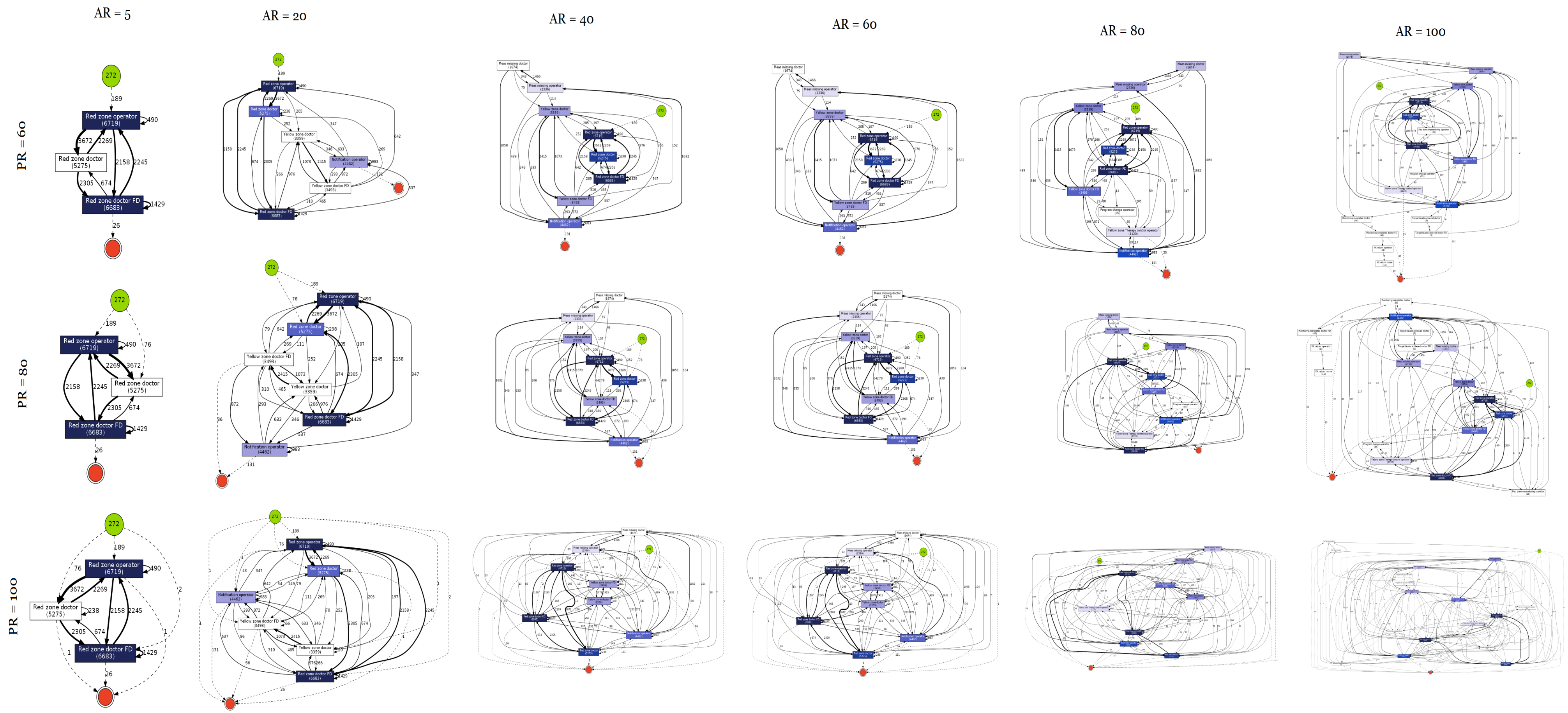


Figure A15. Process model graphs of dataset 1 in R1 and R2.

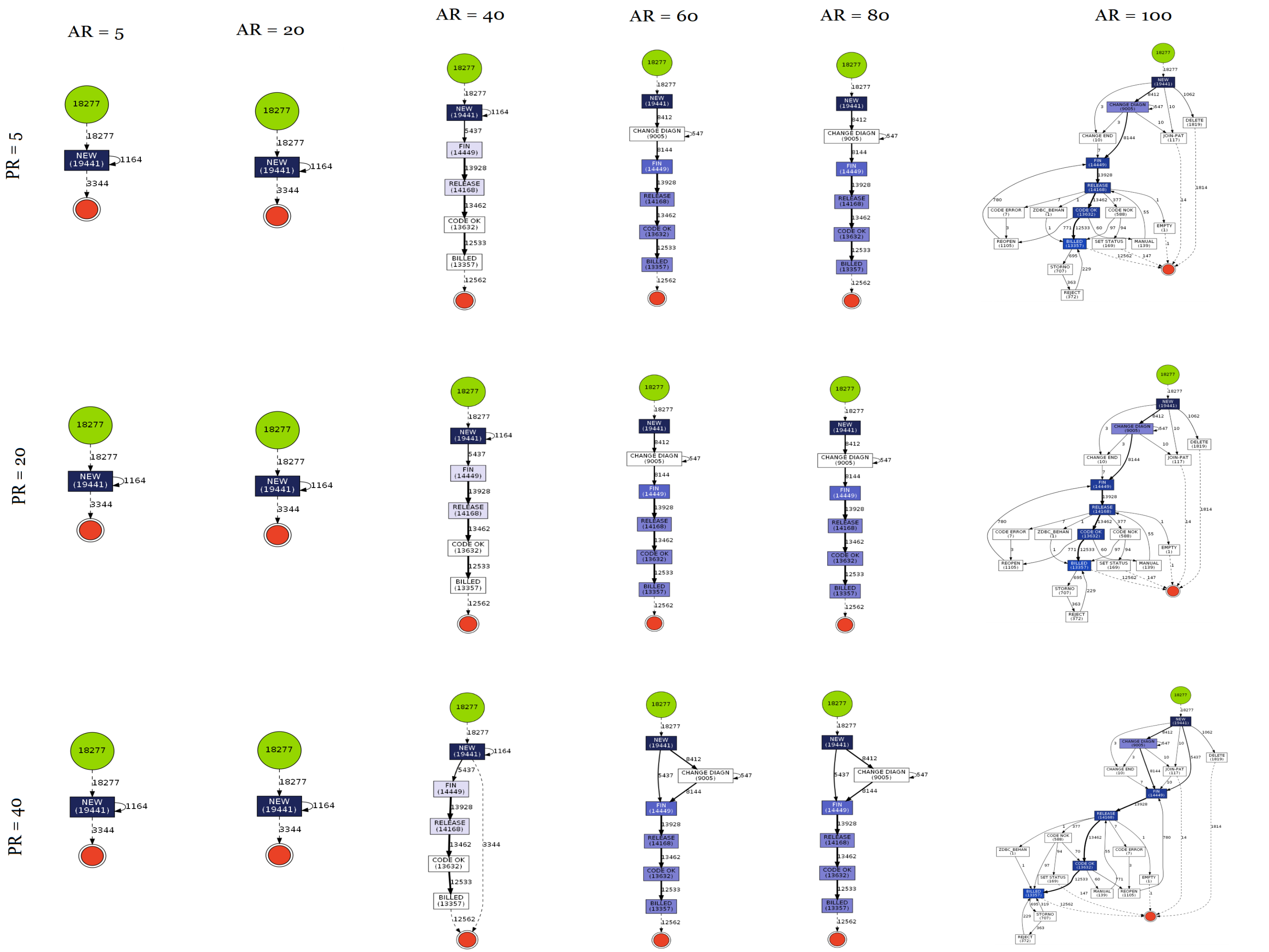


Figure A16. Process model graphs of dataset 3 in R1 and R2.

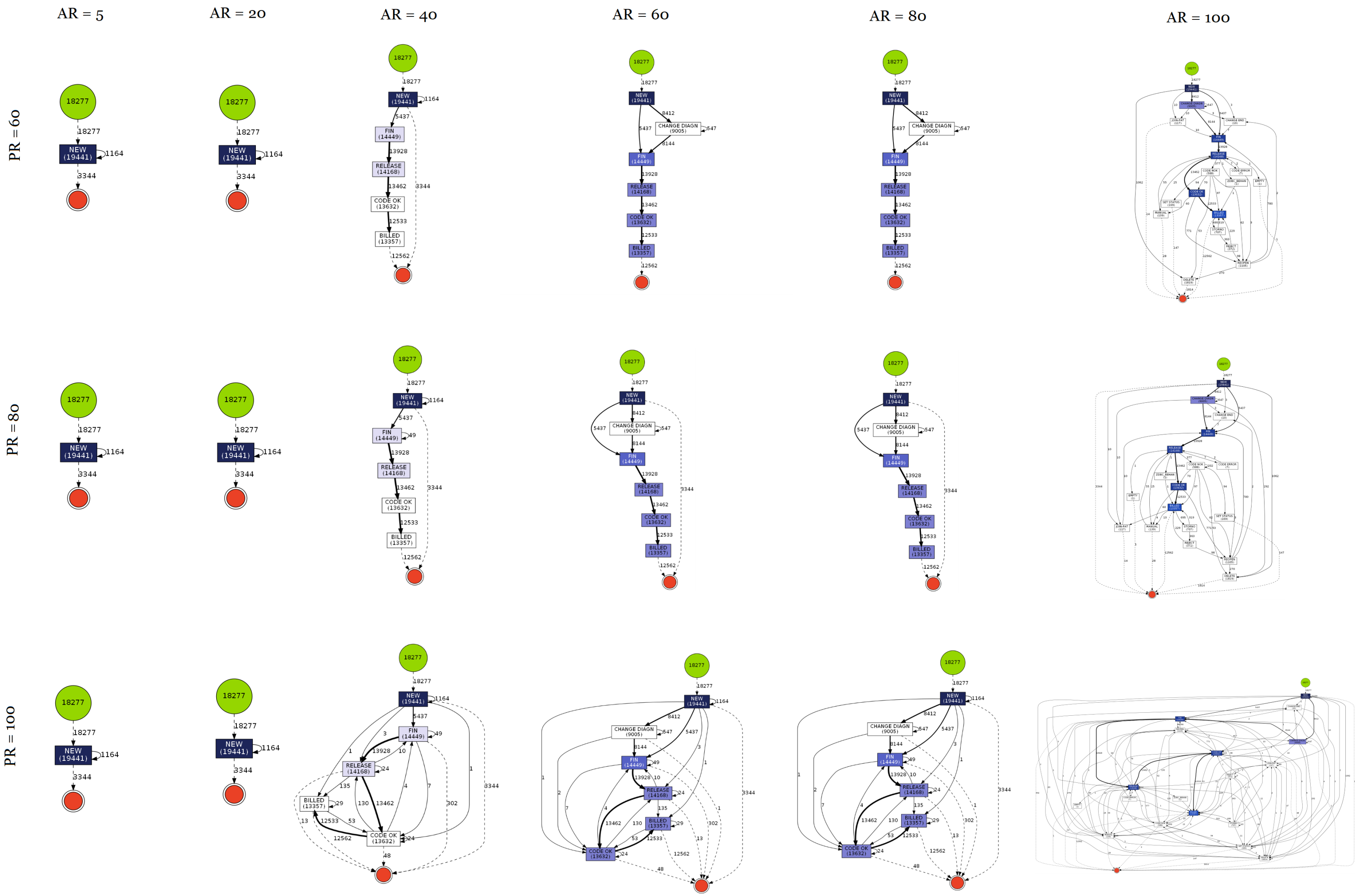


Figure A17. Process model graphs of dataset 3 in R1 and R2.

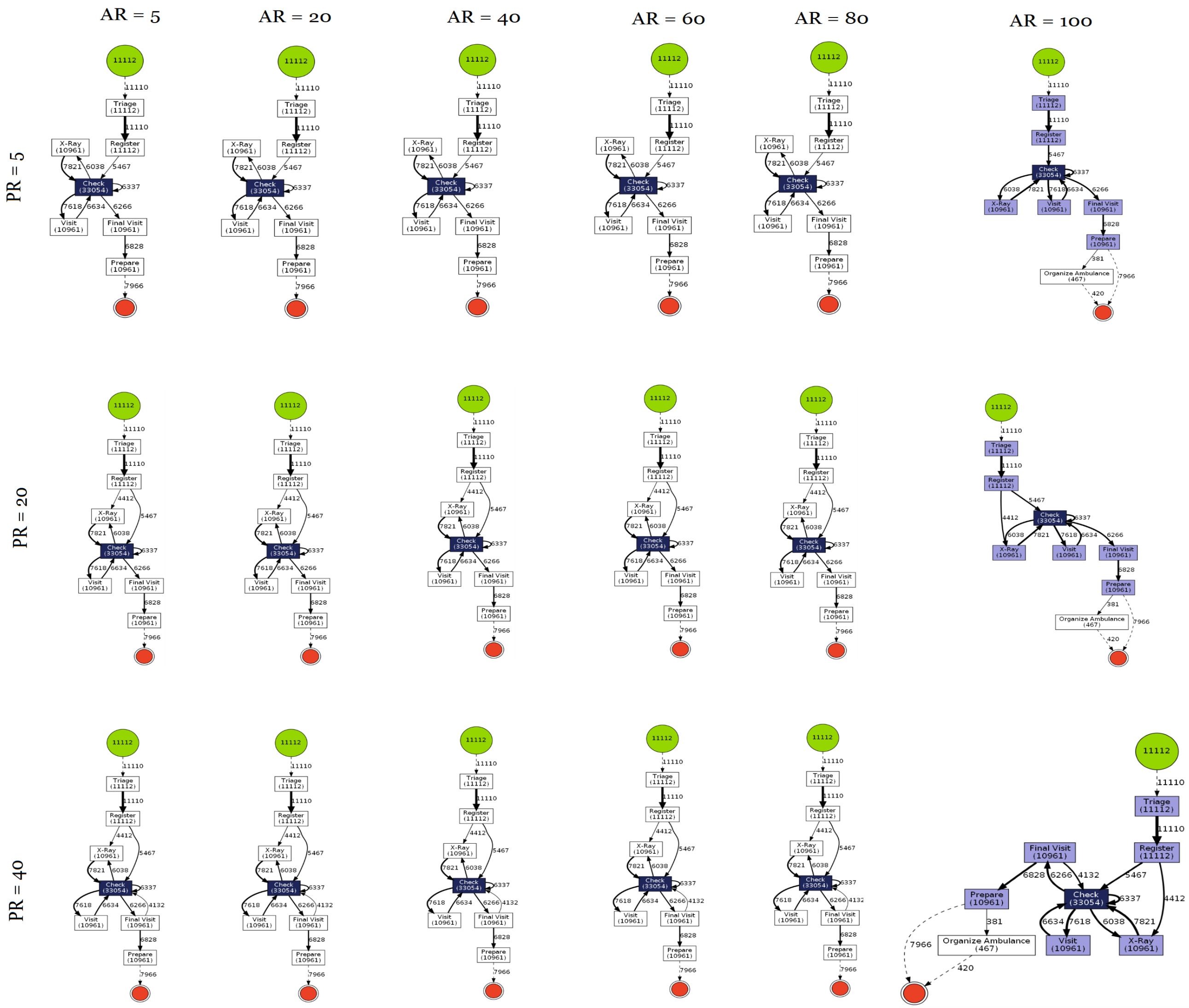


Figure A18. Process model graphs of dataset 5 in R1 and R2.

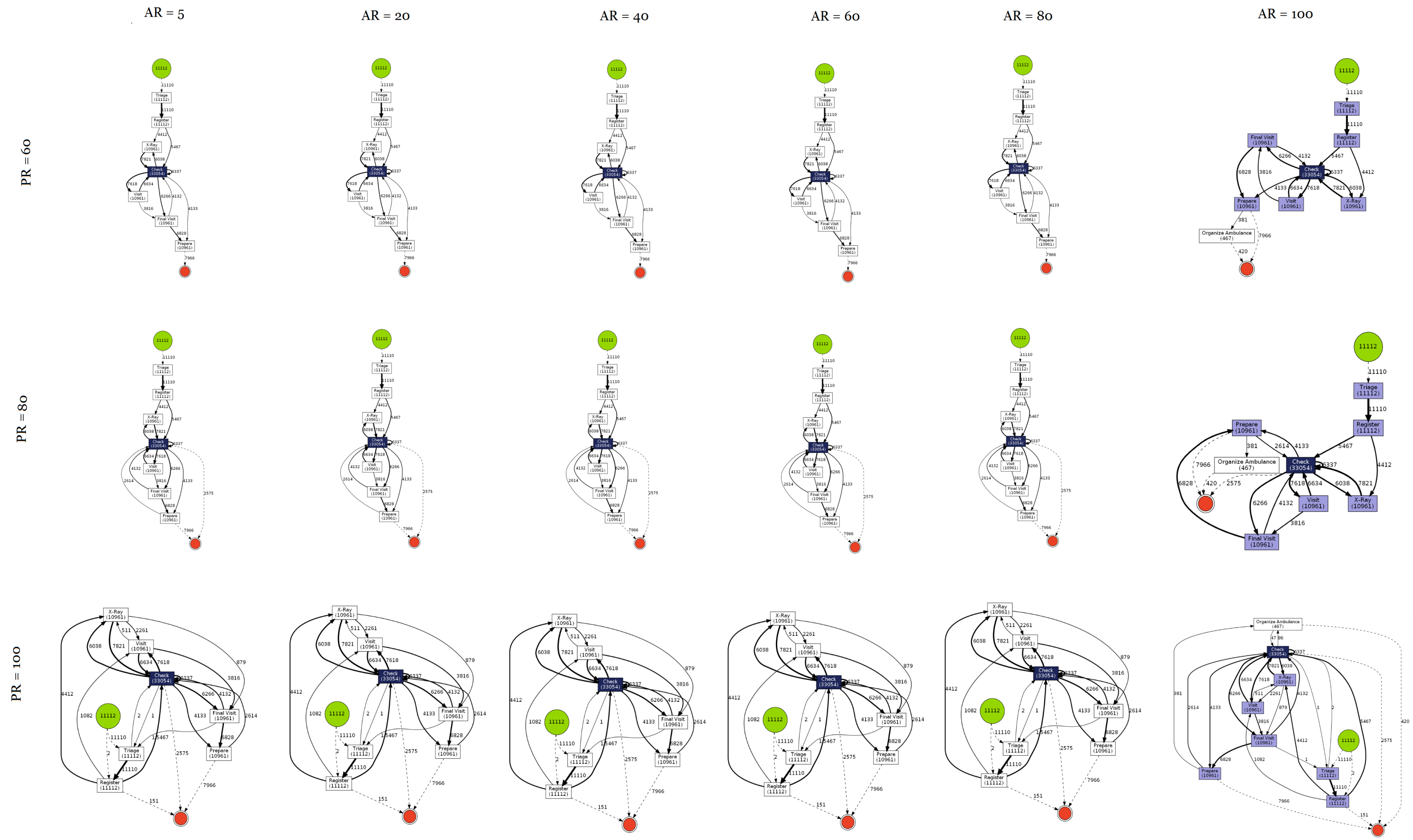


Figure A19. Process model graphs of dataset 5 in R1 and R2.

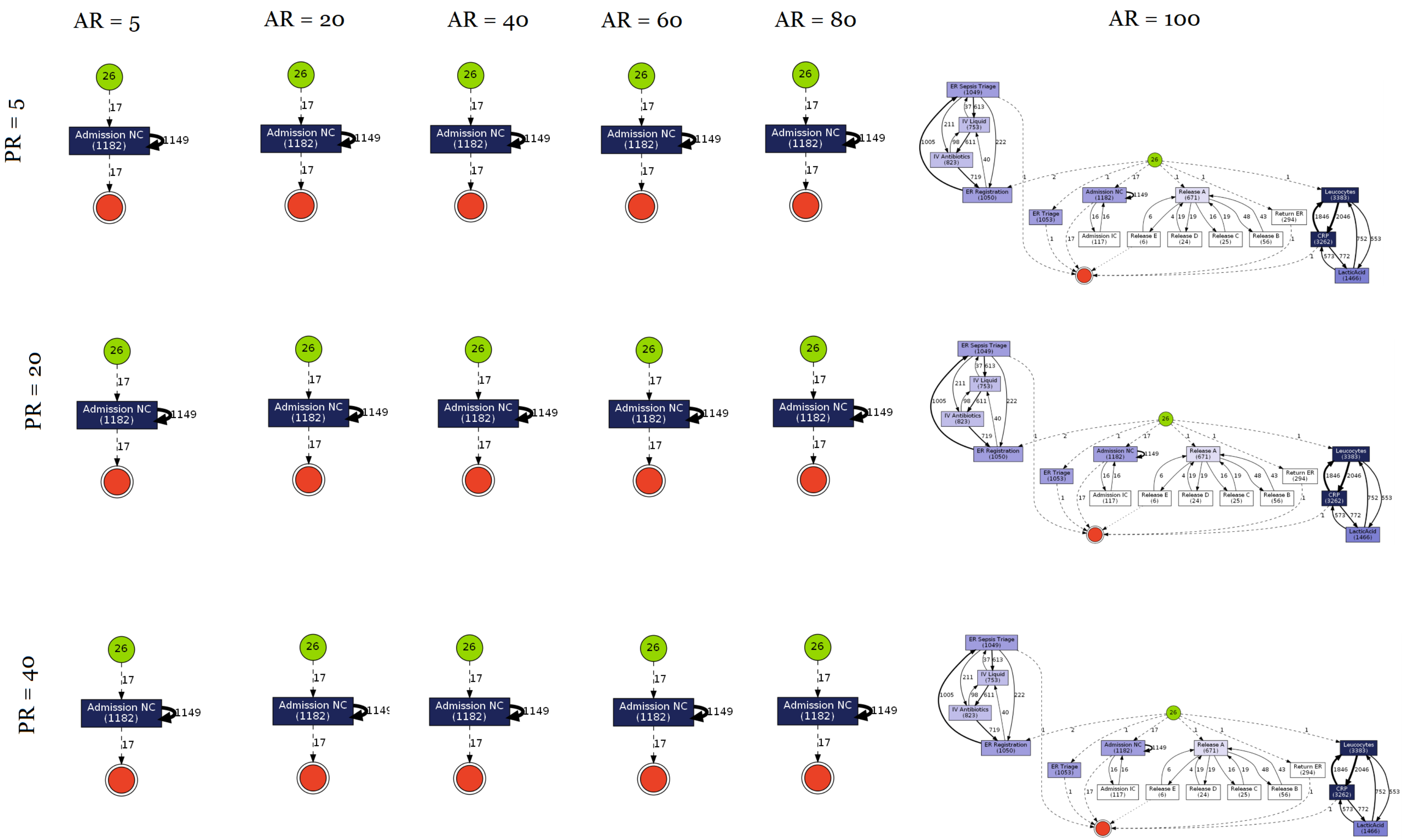


Figure A20. Process model graphs of dataset 7 in R1 and R2.

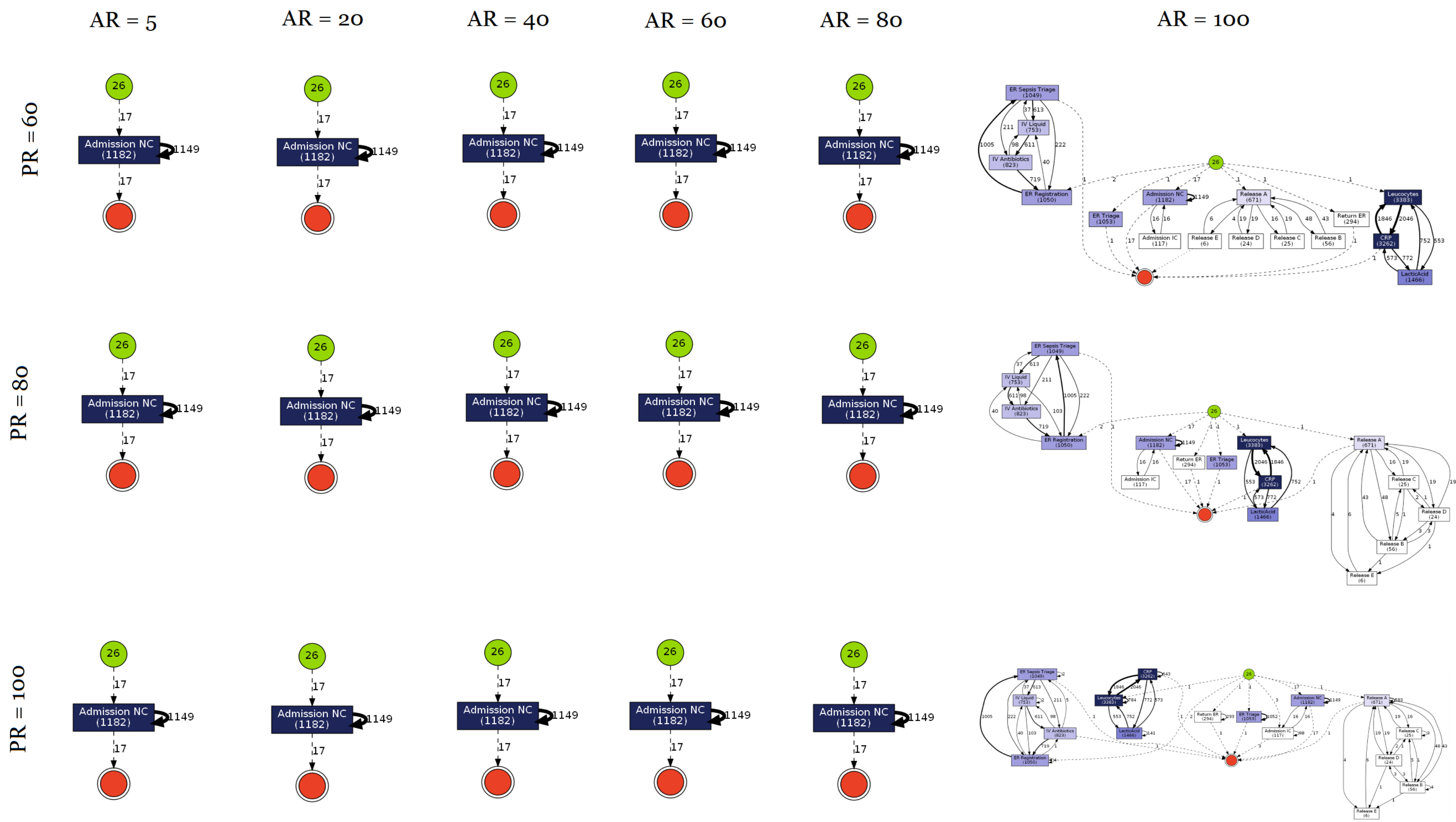


Figure A21. Process model graphs of dataset 7 in R1 and R2.

Appendix C. Taxonomy of Metrics

#	Metric	Year	Auth	Def.	Demo	Type	Src
1	Number of Activities (NOA)	2006	J. Cardoso et al.	Counts the number of activities in a process.	Y	D	[83]
2	Number of Activities and Control flow elements (NOAC)	2006	J. Cardoso et al.	Counts the number of activities and control flow elements in a process.	Y	D	[83]
3	Number of Activities, Joins and Splits (NOAJS)	2006	J. Cardoso et al.	Counts the number of activities, splits and joins in a process. (XOR, OR, AND)	Y	D	[83]
4	Cyclomatic Complexity (CYC)	1976	T.J. McCabe	Indicates the process's control flow complexity by counting number of edges (e) and nodes (n) in the process graph.	Y	D	[105]
5	Control Flow Complexity (CFC)	2005	J. Cardoso	The metric is calculated by adding the CFC's of all split constructs (XOR, OR, AND).	Y	PD	[106, 107]
6	Process Volume (HPC_V)	2006	J. Cardoso et al.	Calculates the volume of a process using number of nodes and edges.	Y	D	[83]
7	Process Length (HPC_N)	2006	J. Cardoso et al.	Calculates a length of the process using number of nodes and edges.	Y	D	[83]
8	Process Difficulty (HPC_D)	2006	J. Cardoso et al.	Calculates a difficulty of the process using number of nodes and edges.	Y	D	[83]
9	Complexity of a Procedure (PC)	1981	S. Henry, D. Kafura	Using the number of calls to/from the module the metric is evaluated as *PC = length * (fan-in * fan-out)^2*.	Y	D	[108]
10	Interface Complexity (IC)	2006	S. Henry, D. Kafura	Using the number of inputs and outputs of an activity, the metric is evaluated as *IC = length * (number of inputs * number of outputs)^2*.	Y	D	[108]
11	Coefficient of Network Complexity (CNC)	2006	Antti M. Latva-Koivisto	Complexity CNC is calculated by dividing the number of arcs by the number of activities, joins and splits. *(No of arcs / (No of activities, joins, and splits))*.	N	D	[1]
12	Restrictiveness estimator (RT)	2006	Antti M. Latva-Koivisto	The metric calculates the number of sequences in the graph.	N	D	[1]
13	Complexity index (CI)	2006	J. Cardoso et al.	The metric performs an algorithmic calculation of the minimal number of node reductions required to reduce the process graph into a single node.	N	D	[83]
14	Cognitive Complexity Metrics (Weights)	2006	V Gruhn, R Laue	Using a pre-defined cognitive weight scale for BPM, the individual weight of the model is generated.	N	PD	[109]
15	Cognitive Functionality Size (CFS)	2003	Shau and WangJ Shao, Y Wang	Evaluated using the cognitive functional size of the model.	Y	S	[110]
16	Extended Cardoso Metrics (EcaM)	2009	KB Lassen, WMP van der Aalst	The metric follows Cardoso approach for petri nets. It works by penalizing direct successor states in the model.	Y	D	[77]
17	Extended Cyclomatric Metrics (EcyM)	2009	KB Lassen, WMP van der Aalst	The behavioural complexity of the graph is calculated by creating and analysing a reachability matrix.	Y	D	[77]
18	Structuredness Metric (SM)	2009	KB Lassen, WMP van der Aalst	The structuredness of a model is a combination of behavioural and syntax complexity together.	Y	PD	[77]

#	Metric	Year	Auth	Def.	Demo	Type	Src
19	Coupling	2015	J Cardoso et al.	Measured by counting the number of interconnections between modules. (If two activities have one ore more common data elements then they are termed as a 'couple').	Y	PD	[111]
20	Cohesion (Activity and Process Cohesion)	2015	J Cardoso et al.	The relativity between elements of a module is evaluated by determining the mean of all activity cohesion values. (i.e., sum of all cohesion values divided by the number of activities).	Y	PD	[111, 112]
21	Modularity	2015	J Cardoso et al.	A black box approach defined by authors. It utilises any visual component of process model graphs to measure overall modularity.	Y	S	[111]
22	Size and Complexity Size	2015	J Cardoso et al.	Direct measurement of the size of the process model using all available features of the model. Similar to LOC metrics.	Y	D	[111]
23	Cross Connectivity Metric (CC)	2008	I Vanderfeesten et al.	Measures the strength of connections between nodes in a model and provides a relation between tightly knit nodes vs weakest link.	Y	D	[113]
24	Separability	2007	J Mendling, G Neumann	Calculated by obtaining a ratio of the number of cut vertices to the number of nodes.	Y	D	[114]
25	Sequentiality	2007	J Mendling, G Neumann	Calculated as the ratio of arcs of a sequence to the total number of arcs.	Y	PD	[114]
26	Structuredness	2007	J Mendling, G Neumann	Calculated as the numerical ratio of the number of nodes in the reduced process graph to the number of nodes in the original graph.	Y	PD	[114]
27	Cyclicity	2007	J Mendling, G Neumann	Calculated as the ratio between the number of nodes in any cycle in the model to the total number of nodes.	Y	PD	[114]
28	Parallelism	2007	J Mendling, G Neumann	The number of concurrent paths obtained after introduction of new control nodes such as AND or OR.	Y	PD	[114]
29	Precision	2016	HT Yang et al.	The number of correctly retrieved relationships divided by the total number of retrieved relationships.	Y	PD	[115]
30	Recall	2016	HT Yang et al.	The number of correctly retrieved relationships divided by the total number of correct relationships possible.	Y	PD	[115]
31	F-Measure			Generic harmonic mean of precision and recall.			
32	Precision at k	2017	WJ Vlietstra et al.	The number of reference compounds found up to rank k, divided by k (where k is the subset of compounds in the process graph).	Y	PD	[116]
33	Recall at k	2017	WJ Vlietstra et al.	The fraction of reference compounds found up to rank k (where k is the subset of compounds in the process graph).	Y	PD	[116]
34	Average Precision	2014	N Shang et al.	The average of the precision values measured at the point at which each correct results is retrieved for one example/trace.	Y	PD	[117]

#	Metric	Year	Auth	Def.	Demo	Type	Src
35	Mean Average Precision	2014	N Shang et al.	The mean of average precision across all examples/samples.	Y	PD	[117]
36	Precision over time	2006	M Yetisgen-Yildiz, W Pratt	Precision calculated at time intervals which coincide with the time of event log data.	Y	PD	[118]
37	Recall over time	2006	M Yetisgen-Yildiz, W Pratt	Recall calculated at time intervals which coincide with the time of event log data.	Y	PD	[118]
38	11-point average interpolated precision	2009	M Yetisgen-Yildiz, W Pratt	Relational evaluation to check how precision changes as recall levels increase for different log based datasets.	Y	S	[119]
39	Area Under Curve			Generic plot of metric operating characteristics for Models.			
40	Accuracy	2015	S Sang et al.	Amount of activity that is not supposed to be observed in the event log (similar to precision).	N	D	[120]
41	Cumulative Gain	2017	WJ Vlietstra et al.	Calculated by dividing the number of the weight points found up to rank k by k.	Y	D	[116]
42	Mean Reciprocal Rank			The mean of the reciprocal rank of the highest ranking correct answer.			
43	Correlation Analysis			Generic approach to relate one metric against another.			
44	Density Metric	2008	J Mendling	A structural metric calculated as a ratio of total number of arcs to the maximum number of arcs.	Y	PD	[84]
45	Imported Coupling of a Process metric (ICP)	2009	W Khlif et al.	The number of messages per subprocess divided by the sequence flows sent by the task of the subprocess or the subprocess itself.	N	PD	[121]
46	Exported Coupling of a Process metric (ECP)	2009	W Khlif et al.	The number of messages per subprocess divided by the sequence flows Received by the task of the subprocess or the subprocess itself.	N	PD	[121]
47	Fan-in/Fan-out metric (FIO)	2006	V Gruhn, R Laue	The count of all incoming and outgoing edges in a module of a graph evaluated using the formulation: $((\textit{fan-in}) \times (\textit{fan-out}))^2$.	N	PD	[122]
48	Diameter	2008	J Mendling	The length of the longest path from a start node to an end node in a process model.	Y	S	[84]
49	Number of nodes metric (Sn(G))	2008	J Mendling	Number of nodes in a process model graph.	Y	S	[84]
50	Degree of connectors	2008	J Mendling	The average number of nodes a connector is connected to.	Y	S	[84]
51	Sequentiality metric (S(G)) /Sequentiality ratio	2008	J Mendling	The ratio between the number of arcs between none connector nodes divided by the total number of arcs.	Y	D	[84]
52	Max depth	2008	J Mendling	The maximum depth of all nodes.	Y	D	[84]
53	Concurrency	2008	J Mendling	The sum of the output—degree of AND-joins and OR-joins minus one.	Y	D	[84]
54	Heterogeneity	2008	J Mendling	The entropy of the model over different connector types.	Y	D	[84]
55	Cognitive complexity measure (W)	2006	V Gruhn, R Laue	A cognitive weight that measures the effort needed for comprehending the model.	Y	D	[109]

#	Metric	Year	Auth	Def.	Demo	Type	Src
56	Fitness	2012	JCAM Buijs et al.	The ability of a model to reproduce the behaviour contained in a log.	Y	S	[86]
57	Generalization K fold cross validation	1995	R Kohavi	The ability of an automated discovery algorithm to discover process models that generate traces that are not present in the log, but can be produced by the business process during operation.	N	S	[123]
58	Generalization	2012	JCAM Buijs et al.	The extent to which the resulting model will be able to reproduce future behaviour of the process.	Y	PD	[86]
59	Alignment based Precision	2012	JCAM Buijs et al.	The ability of a model to generate only the behaviour found in the log or the amount of behaviour that is present only in the log and not made in the model.	Y	S	[86]
60	Simplicity	2012	JCAM Buijs et al.	The number of activities in the log is used to measure the perceived complexity of the model.	Y	S	[86]
61	Soundness	1997	WMP Van der Aalst	Evaluates the behavioural quality of a model by checking for incomplete criteria such as option to complete, proper completion and absence of dead transitions.	Y	S	[124]
62	Number of sub processes	2018	N Wang et al.	The total number of subprocesses in the model.	Y	S	[125]
63	Place/Transition degree (P/T – CD)	2018	N Wang et al.	The weighted sum of average number of arcs per transition to the average number of arcs per place.	Y	S	[125]
64	Cyclomatic Number (CN)	2018	N Wang et al.	The number of linearly independent paths in a process model where directions of the arcs are ignored (A measure of branching).	Y	D	[125]
65	Average Connector Degree (ACD)	2018	N Wang et al.	Measures the average number of connecting nodes by calculating the average count of incoming and outgoing arcs of places/ transitions per connector.	Y	PD	[125]
66	Average number of activities per sub process	2018	N Wang et al.	The metric is evaluated as the average number of activities per subprocess for different levels of process model abstraction.	Y	PD	[125]
67	Number of event classes and variants	2018	N Wang et al.	A count of unique process variants and unique activity classes.	Y	PD	[125]
68	Split-join ratio	2006	V Gruhn, R Laue	Evaluates the number of incoming and outgoing elements at splits (XOR, OR, AND).	Y	D	[122]
69	Nesting Depth (ND)	2006	V Gruhn, R Laue	The number of decisions in the control flow necessary to perform an activity.	Y	PD	[122]
70	Cognitive Complexity (CC)	2006	Y Wang, J Shao	Using derived weights for control structures, It is a measure of difficulty of understandability of a process model.	Y	S	[126]
71	Average Gateway Degree (AGD)	2012	L Sánchez-González et al.	The average number of incoming and outgoing edges of gateway nodes in a process model.	Y	S	[127]
72	Maximum Gateway Degree (MGD)	2012	L Sánchez-González et al.	The maximum sum of incoming and outgoing edges of the gateways.	Y	S	[127]

#	Metric	Year	Auth	Def.	Demo	Type	Src
73	Gateway Mismatch (GM)	2012	L Sánchez-González et al.	The sum of gateway pairs that do not have match with each other.	Y	S	[127]
74	Gateway Heterogeneity (GH)	2012	L Sánchez-González et al.	The extent to which different types of gateways are used in the model.	Y	S	[127]
75	Total Number of Sequence Flows (TNSF)	2006	E Rolón et al.	Total number of sequence flows in a graph.	Y	D	[128, 129]
76	Total Number of Events (TNE)	2006	E Rolón et al. and L Sánchez-González et al.	Total number of events in the model.	Y	D	[120, 128]
77	Total Number of Gateways (TNG)	2006	E Rolón et al.	Total number of gateways in the model.	Y	D	[128]
78	Number of Sequence Flows from Events (NSFE)	2006	E Rolón et al.	Number of sequence flows incoming from an event.	Y	D	[128]
79	Number of Message Flows (NMF)	2006	E Rolón et al.	Number of message flows between participants in the process.	Y	D	[128]
80	Number of Sequence Flows from Gateways (NSFG)	2006	E Rolón et al.	Number of sequence flows incoming from gateway.	Y	D	[128]
81	Connectivity Level between Pools (CLP)	2006	E Rolón et al.	Connectivity level between pools of activities.	Y	D	[128]
82	Total Number of Data Objects (TNDO)	2006	E Rolón et al.	Total number of data objects in the process model.	Y	D	[128]
83	Number of Inclusive Decisions (NID)	2006	E Rolón et al.	Indicates the number of points of inclusive decision, and merging of the model.	Y	D	[128]
84	Number of Parallel Forking (NPF)	2006	E Rolón et al.	Indicates the number of points of parallel forking and joining of the process.	Y	D	[128]
85	Number of Pools (NP)	2006	E Rolón et al.	Number of pools in the process.	Y	D	[128]
86	Number of Complex Decisions (NCD)	2006	E Rolón et al.	Indicates the number of points of complex decision merging of the model.	Y	D	[128]
87	Number of Exclusive gateways based on Data (NEDDB)	2006	E Rolón et al.	Indicates the number of points of exclusive decision and merging based on data of the model.	Y	D	[128]
88	Number of Exclusive gateways based on Events (NEDEB)	2006	E Rolón et al.	Indicates the number of points of exclusive decision and merging based on events of the model.	Y	D	[128]
89	Number of Intermediate Message Events (NIMsE)	2006	E Rolón et al.	Number of intermediate messages between events.	Y	D	[128]

#	Metric	Year	Auth	Def.	Demo	Type	Src
90	Number of End Message Events (NEMsE)	2006	E Rolón et al.	Number of end message events.	Y	D	[128]
91	Total Number of Collapsed processes (TNCS)	2006	E Rolón et al.	Total number of collapsed sub-process of the model.	Y	D	[128]
92	Connectivity Level between Activities (CLA)	2006	E Rolón et al.	Connectivity level between activities.	Y	D	[128]
93	Anti-patterns	2000	J Paakki et al.	Commonly occurring solutions to a problem that are known to have negative consequences is evaluated.	Y	S	[18, 130]
94	Knot count	2007	V Gruhn et al. and MR Woodward et al.	A measure of the number of paths associated when transfer of control intersect (overlap).	Y	PD	[18, 131]
95	Log-Based Complexity (LBC)	2007	J Cardoso	The number of unique log traces that can be generated from the execution of a workflow.	Y	S	[132]
96	Average Vertex degree ((A)VG)	2008	J Mendling	Summarises whether vertices are connected to many or to few other vertices, where degree d(v) of a vertex is the number of edges that are connected to it.	Y	S	[84]
97	Quantity of Decisions to be made per pool/participant (CUDP)	2010	N Debnath et al.	Evaluates the quantity of decision nodes inside a pool in the model.	Y	S	[133]
98	Quantity of tasks executed in a specific pool/participant (CTP)	2010	N Debnath et al.	Evaluates the the load level for each pool.	Y	S	[133]
99	Quantity of tasks of a swim-lane of a pool (CTSP)	2010	N Debnath et al.	Evaluates the organization and distribution of tasks inside a pool.	Y	S	[133]
100	Proportion of task distribution per Participant (PTP)	2010	N Debnath et al.	Evaluates the proportion of tasks for one task in relation to the total of process tasks.	Y	S	[133]
101	Proportion of tasks per swim-lane of a Specific Pool (PTSP)	2010	N Debnath et al.	Calculated as the proportion of tasks per actor (swim-lane) of a specific participant (pool).	Y	S	[133]
102	Quantity of Subprocesses per pool (NSBPart)	2010	N Debnath et al.	The number of sub-processes per participant.	Y	S	[133]
103	Quantity of Message Flows between two pools (NFPart)	2010	N Debnath et al.	The number of messages (Flowing) between two participants.	Y	S	[133]
104	Durfee Square Metric (DSP)	2012	K Kluza, GJ Nalepa	Calculates the relation between occurrence, frequency and threshold of an element.	Y	D	[134]
105	Perfect Square Metric (PSM)	2012	K Kluza, GJ Nalepa	It is Pth (unique) largest number such that the top element occurs at least (P×2) times.	Y	D	[134]
106	Structural complexity by Cheng	2008	Cheng, Chen-Yang	The expected amount of information required for defining the state of the process flow.	Y	S	[135]
107	Interaction Complexity by Cheng	2008	Cheng, Chen-Yang	The relation between average information in the model and complexity of the same model.	Y	S	[135]
108	Usability Complexity by Cheng	2008	Cheng, Chen-Yang	The relation between the number of interactions in the model vs the operators required to complete a task in consideration.	Y	S	[135]
109	Total Operational Complexity by Cheng	2008	Cheng, Chen-Yang	The euclidean norm of the structural, interactional and usability complexities used according to their weightage.	Y	S	[135]

#	Metric	Year	Auth	Def.	Demo	Type	Src
110	GQM-based Complexity Metrics	2008	AAA Ghani et al.	Used to asses the understandability and maintainability of the process model by designing a set of questions aimed to fulfil the goal and adapting the metrics to them.	Y	D	[136]
111	Gateway Complexity Indicator (GCI)	2012	L Sánchez-González et al.	It is the weighted sum of (CFC, GM, GH, AGD, MGD, TNG).	Y	D	[20]
112	Trace based L2M precision and recall	2019	AF Syring et al.	Recall is defined as the number of traces occurring in both event log and model divided by the number of traces in the log, The precision is defined as the number of traces occurring in both log and model divided by the number of traces in the model.	N	PD	[137]
113	Frequency based L2M recall	2019	AF Syring et al.	The number of times a unique trace occurs in both event log and model divided by the length of the trace.	N	PD	[137]
114	Causal footprint recall	2019, 2004	AF Syring et al. and WMP Van der Aalst et al.	The causal dependency between two activities. (If activity X is followed by Y but Y is never followed by X then, there is a causal dependency between X and Y.)	Y	D	[3, 137]
115	Token replay recall	2019, 2008	AF Syring et al. and A Rozinat et al.	The recall is calculated by replaying the log on the model and counting the mismatches as missing and remaining tokens.	Y	D	[3, 137]
116	Alignment recall	2012	WMP Van der Aalst et al.	The metric maps steps taken in the event log to that of the model and tracks the deviations occurring between them during replay of the logs.	Y	D	[138]
117	Behavioural recall	2009	S Goedertier et al.	The metric is determined using true positive and false negative conditions between event log and process model (i.e. Transitional state of activities).	Y	PD	[139]
118	Projected recall	2018	SJJ Leemans et al.	Calculated by projecting the event log and model on all possible subsets of activities of size k and solving the fraction of behaviour allowed by the minimal log-automaton, then divided by the allowed minimal model-automaton per projection.	Y	PD	[140]
119	Continued parsing measure	2006	A Weijters et al.	Calculated by counting the number of input and output activities, in active and inactive state w.r.t the model and the event log.	Y	PD	[141]
120	Eigenvalue recall	2018	A Polyvyanyy et al.	Calculated by evaluating the relational eigenvalues of event log and metric and setting them in relation.	Y	PD	[142]
121	Simple behavioural appropriateness	2008	A Rozinat et al.	The mean number of enabled/used transitions for each unique trace in relation to the visible traces in the process model.	Y	PD	[143]
122	Advanced behavioural appropriateness	2008	A Rozinat et al.	Calculated by describing the relation between event log and process model by analysing whether activities follow each other or precede each other.	Y	PD	[143]
123	ETC-one/ETC-rep	2010	J Munoz-Gama et al.	Evaluates the states of the model visited by the event log. For each, state the precision, which is calculated as the weighted sum of non escaping edges to the total edges.	Y	PD	[144]
124	Behavioural specificity (precM) and Behavioural precision	2009	S Goedertier et al.	Calculated by generating a confusion matrix of true positive, false positive and true negative relations between model and event log.	Y	PD	[139]
125	Weighted negative event precision	2013	SKLM vanden Broucke et al.	Calculated by generating matching subsets of preceding events in the log and finding their occurrence frequency.	Y	PD	[145]

#	Metric	Year	Auth	Def.	Demo	Type	Src
126	Projected precision	2018	SJJ Leemans et al.	Calculated by analysing the conjunction of behaviour between model and event log. The final numeric is attained by making subsets of activities and averaging against the total number of subsets.	Y	PD	[140]
127	Anti-alignment precision	2016	BF van Dongenm et al.	The anti-alignment of each trace averaged by the total number of traces, where a single trace removed from the log makes it impossible for the rest of the log and model, to be precise.	Y	PD	[146]
128	Eigenvalue precision	2018	A Polyvyanyy et al.	Calculated as the relational behavioural eigenvalues of log and metric when in relation.	Y	PD	[142]
129	Weighted negative event generalization	2013	SKLM vanden Broucke et al.	The weightage of events that could be replayed without errors that confirm the model is general.	Y	PD	[145]
130	Anti-alignment generalization	2016	BF van Dongenm et al.	The maximum distance between the states visited by the log and the states visited by the anti-alignment log (i.e., the subset of log without a trace).	Y	PD	[146]
131	Completeness	2009	C Batini et al.	Calculated as the number of NOT null values divided by the total number of values in event log or model (values can be activities or edges).	N	D	[147]
132	Consistency	2009	C Batini et al.	Calculated as the number of consistent values divided by the number of total values (values can be activities or edges).	N	D	[147]
133	Uniqueness	2009	C Batini et al.	Calculated as the number of duplicates in the log or model.	N	D	[147]
134	Appropriate amount of data	2009	C Batini et al.	The number of data units needed to represent a trace or log divided by the number of data units provided in the model.	N	D	[147]
135	Tasks	2008	H Reijers, J Mendling	Total number of tasks in the model or event log.	Y	PD	[148]
136	Nodes	2008	H Reijers, J Mendling	Total number of nodes in the process model.	Y	PD	[148]
137	Arcs	2008	H Reijers, J Mendling	Total number of arcs in the model.	Y	PD	[148]
138	Subproc	2008	H Reijers, J Mendling	Total number of subprocesses in model.	Y	PD	[148]
139	To	1992	LG Soo, Y Jung-Mo	Average number of outgoing edges from transitions (tasks).	Y	PD	[149]
140	Po	1992	LG Soo, Y Jung-Mo	Average number of outgoing edges from places (milestones).	Y	PD	[149]

References

1. Latva-Koivisto, A.M. *Finding a Complexity Measure for Business Process Models*; Research Report; Helsinki University of Technology: Espoo, Finland, 13 February 2001.
2. Rozinat, A.; De Medeiros, A.A.; Günther, C.W.; Weijters, A.; Van der Aalst, W.M. *Towards an Evaluation Framework for Process Mining Algorithms*; BPM Center Report BPM-07-06; BPMcenter. org; Eindhoven University of Technology: Eindhoven, The Netherlands, 2007; Volume 123, p. 142.
3. Van der Aalst, W.; Weijters, T.; Maruster, L. Workflow mining: Discovering process models from event logs. *IEEE Trans. Knowl. Data Eng.* **2004**, *16*, 1128–1142. [CrossRef]
4. Kueng, P.; Kawalek, P. Goal-based business process models: Creation and evaluation. *Bus. Process. Manag. J.* **1997**, *3*, 17–38. [CrossRef]
5. Hommes, B.J.; Van Reijswoud, V. Assessing the quality of business process modelling techniques. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 7 January 2000; p. 10.
6. Rozinat, A.; Medeiros, A.; Günther, C.; Weijters, A.; Aalst, W. The Need for a Process Mining Evaluation Framework in Research and Practice In *Business Process Management Workshops: BPM 2007 International Workshops, BPI, BPD, CBP, ProHealth, RefMod, semantics4ws, Brisbane, Australia, 24 September 2007*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4928, pp. 84–89. [CrossRef]
7. Turner, C.J.; Tiwari, A.; Olaiya, R.; Xu, Y. Process mining: From theory to practice. *Bus. Process. Manag. J.* **2012**, *18*, 493–512. [CrossRef]
8. Born, A.; Dorsan, N.; Therese Lalor, U.; Özkan, D.; Signore, M. *Recent developments in the Generic Statistical Business Process Model: Revisions and Quality Indicators*; European Conference on Quality in Official Statistics: Krakow, Poland, 2018.
9. Heidari, F.; Loucopoulos, P. Quality evaluation framework (QEF): Modeling and evaluating quality of business processes. *Int. J. Account. Inf. Syst.* **2014**, *15*, 193–223. [CrossRef]
10. Dani, V.S.; Freitas, C.M.D.S.; Thom, L.H. Ten years of visualization of business process models: A systematic literature review. *Comput. Stand. Interfaces* **2019**, *66*, 103347. [CrossRef]
11. Van Der Aalst, W. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*; Springer: Heidelberg/Berlin, Germany, 2011; Volume 2.
12. Van Der Aalst, W. *Process Mining: Data Science in Action*; Springer: Heidelberg/Berlin, Germany, 2016; Volume 2.
13. Reinkemeyer, L. *Process Mining in Action Principles, Use Cases and Outlook: Principles, Use Cases and Outlook*; Springer: Cham, Switzerland, 2020. [CrossRef]
14. Günther, C.W. *Process Mining in Flexible Environments*; Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2009. [CrossRef]
15. Burattin, A. *Process Mining Techniques in Business Environments*; Lecture Notes in Business Information Processing; Springer: Cham, Switzerland, 2015; Volume 207.
16. Fernandez-Llatas, C. *Interactive Process Mining in Healthcare*; Springer: Cham, Switzerland, 2021. [CrossRef]
17. Van der Aalst, W.M.; Weijters, A.J. Process mining: A research agenda. *Comput. Ind.* **2004**, *53*, 231–244. [CrossRef]
18. Gruhn, V.; Laue, R. Approaches for business process model complexity metrics. In *Technologies for Business Information Systems*; Springer: Dordrecht, The Netherlands, 2007; pp. 13–24.
19. Mendling, J.; Recker, J.; Reijers, H.A. *Process Modeling Quality: A Framework and Research Agenda*; BPM Center Report BPM-09-02, BPM Center Report. org; Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2009. Available online: https://www.win.tue.nl/~hreijers/H.A.%20Reijers%20Bestanden/TechnicalReport.pdf (accessed on 30 April 2023).
20. Sánchez-González, L.; García, F.; Ruiz, F.; Mendling, J. Quality indicators for business process models from a gateway complexity perspective. *Inf. Softw. Technol.* **2012**, *54*, 1159–1174. [CrossRef]
21. Dijkman, R.; Dumas, M.; Van Dongen, B.; Käärik, R.; Mendling, J. Similarity of business process models: Metrics and evaluation. *Inf. Syst.* **2011**, *36*, 498–516. [CrossRef]
22. Augusto, A.; Armas-Cervantes, A.; Conforti, R.; Dumas, M.; La Rosa, M. Measuring fitness and precision of automatically discovered process models: A principled and scalable approach. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 1870–1888. [CrossRef]
23. Sadowska, M. An approach to assessing the quality of business process models expressed in BPMN. *E-Inform. Softw. Eng. J.* **2015**, *9*, 57–77.
24. Wynn, D.C.; Clarkson, P.J. Process models in design and development. *Res. Eng. Des.* **2018**, *29*, 161–202. [CrossRef]
25. Martin, N.; De Weerdt, J.; Fernández-Llatas, C.; Gal, A.; Gatta, R.; Ibáñez, G.; Johnson, O.; Mannhardt, F.; Marco-Ruiz, L.; Mertens, S.; et al. Recommendations for enhancing the usability and understandability of process mining in healthcare. *Artif. Intell. Med.* **2020**, *109*, 101962. [CrossRef]
26. Kubrak, K.; Milani, F.; Nolte, A. Process mining for process improvement-an evaluation of analysis practices. In Proceedings of the Research Challenges in Information Science: 16th International Conference, RCIS 2022, Barcelona, Spain, 17–20 May 2022; Springer: Cham, Switzerland, 2022; pp. 214–230.
27. SáNchez-GonzáLez, L.; GarcíA, F.; Ruiz, F.; Piattini, M. Towards a quality framework for business process models. *Int. J. Coop. Inf. Syst.* **2013**, *22*, 1350003. [CrossRef]
28. Overhage, S.; Birkmeier, D.Q.; Schlauderer, S. Quality marks, metrics, and measurement procedures for business process models: The 3QM-framework. *Bus. Inf. Syst. Eng.* **2012**, *4*, 229–246. [CrossRef]

29. Moreno-Mones de Oca, I.; Snoeck, M. *Pragmatic Guidelines for Business Process Modeling*; FEB Research Report KBI_1509; KU Leuven, Faculty of Economics and Business (FEB), Department of Decision Sciences and Information Management: Leuven, Belgium, 2015.
30. Mendling, J.; Reijers, H.A.; van der Aalst, W.M. Seven process modeling guidelines (7PMG). *Inf. Softw. Technol.* **2010**, *52*, 127–136. [CrossRef]
31. Avila, D.T. Process Modeling Guidelines: Systematic Literature Review and Experiment. Masters Dissertation, Federal University of Rio Grande do Sul, Farroupilha, Brazil, 2018. Available online: https://www.lume.ufrgs.br/handle/10183/175028 (accessed on 30 April 2023).
32. Kahloun, F.; Ghannouchi, S.A. Improvement of quality for business process modeling driven by guidelines. *Procedia Comput. Sci.* **2018**, *126*, 39–48. [CrossRef]
33. Muketha, G.M.; Abd Ghani, A.; Selamat, M.; Atan, R. *A Survey of Business Processes Complexity Metrics*; Asian Network for Scientific Information: Faisalabad, Pakistan, 2010. [CrossRef]
34. John, T.; Kundisch, D.; Szopinski, D. Visual Languages for Modeling Business Models: A Critical Review and Future Research Directions. In Proceedings of the ICIS, Seoul, Republic of Korea, 10–13 December 2017.
35. Jun, G.T.; Ward, J.; Morris, Z.; Clarkson, J. Health care process modelling: Which method when? *Int. J. Qual. Health Care* **2009**, *21*, 214–224. [CrossRef]
36. La Rosa, M.; Wohed, P.; Mendling, J.; Ter Hofstede, A.H.; Reijers, H.A.; van der Aalst, W.M. Managing process model complexity via abstract syntax modifications. *IEEE Trans. Ind. Inform.* **2011**, *7*, 614–629. [CrossRef]
37. La Rosa, M.; ter Hofstede, A.H.M.; Wohed, P.; Reijers, H.A.; Mendling, J.; van der Aalst, W.M.P. Managing Process Model Complexity via Concrete Syntax Modifications. *IEEE Trans. Ind. Inform.* **2011**, *7*, 255–265. [CrossRef]
38. Kluza, K. Measuring complexity of business process models integrated with rules. In Proceedings of the Artificial Intelligence and Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, 14–18 June 2015; Part II 14; Springer: Cham, Switzerland, 2015; pp. 649–659.
39. Augusto, A.; Conforti, R.; Dumas, M.; La Rosa, M.; Maggi, F.M.; Marrella, A.; Mecella, M.; Soo, A. Automated discovery of process models from event logs: Review and benchmark. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 686–705. [CrossRef]
40. Ghasemi, M.; Amyot, D. Process mining in healthcare: A systematised literature review. *Int. J. Electron. Healthc.* **2016**, *9*, 60–88. [CrossRef]
41. dos Santos Garcia, C.; Meincheim, A.; Junior, E.R.F.; Dallagassa, M.R.; Sato, D.M.V.; Carvalho, D.R.; Santos, E.A.P.; Scalabrin, E.E. Process mining techniques and applications—A systematic mapping study. *Expert Syst. Appl.* **2019**, *133*, 260–295. [CrossRef]
42. Yahya, F.; Boukadi, K.; Ben-Abdallah, H. Improving the quality of Business Process Models: Lesson learned from the State of the Art. *Bus. Process. Manag. J.* **2019**, *25*, 1357–1376. [CrossRef]
43. *ISO/IEC 25010*; 2011 Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—System and Software Quality Models. International Organization for Standardization: Geneva, Switzerland, 2013.
44. Carmona, J.; van Dongen, B.; Solti, A.; Weidlich, M. *Conformance Checking*; Springer: Cham, Switzerland, 2018.
45. Munoz-Gama, J. *Conformance Checking and Diagnosis in Process Mining*; Springer: Cham, Switzerland, 2016.
46. Rozinat, A. *Process Mining: Conformance and Extension*; Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2010. [CrossRef]
47. Fani Sani, M.; Boltenhagen, M.; van der Aalst, W. Prototype selection using clustering and conformance metrics for process discovery. In Proceedings of the Business Process Management Workshops: BPM 2020 International Workshops, Seville, Spain, 13–18 September 2020; Revised Selected Papers 18; Springer: Cham, Switzerland, 2020; pp. 281–294.
48. De Meyer, P.; Claes, J. An overview of process model quality literature-the comprehensive process model quality framework. *arXiv* **2018**, arXiv:1808.07930.
49. Becker, J.; Rosemann, M.; Von Uthmann, C. Guidelines of business process modeling. In *Business Process Management: Models, Techniques, and Empirical Studies*; Springer: Heidelberg/Berlin, Germany, 2002; pp. 30–49.
50. Corradini, F.; Ferrari, A.; Fornari, F.; Gnesi, S.; Polini, A.; Re, B.; Spagnolo, O. *Quality Assessment Strategy: Applying Business Process Modelling Understandability Guidelines*; University of Camerino: Camerino, Italy, 2017.
51. Krogstie, J.; Sindre, G.; Jørgensen, H. Process models representing knowledge for action: A revised quality framework. *Eur. J. Inf. Syst.* **2006**, *15*, 91–102. [CrossRef]
52. Lindland, O.; Sindre, G.; Solvberg, A. Understanding quality in conceptual modeling. *IEEE Softw.* **1994**, *11*, 42–49. [CrossRef]
53. Nelson, H.J.; Poels, G.; Genero, M.; Piattini, M. A conceptual modeling quality framework. *Softw. Qual. J.* **2012**, *20*, 201–228. [CrossRef]
54. Reijers, H.A.; Mendling, J.; Recker, J. Business process quality management. In *Handbook on Business Process Management 1: Introduction, Methods, and Information Systems*; Springer: Heidelberg/Berlin, Germany, 2015; pp. 167–185.
55. Van Mersbergen, M. *A Framework for Business Process Model Quality and an Evaluation of Model Characteristics as Predictors for Quality*; Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2013.
56. van der Aalst, W.M.; Carmona, J. *Process Mining Handbook*; Springer: Cham, Switzerland, 2022.
57. ProFIT: Process Flow Investigation Tool-ITMO University. Available online: https://github.com/itmo-escience/ProFIT (accessed on 30 April 2023).

58. PM4PY: State of the Art Process Mining Toolkit-Institute for Applied Technology Fraunhofer. Available online: https://pm4py.fit.fraunhofer.de/ (accessed on 30 April 2023).
59. Elkhovskaya, L.O.; Kshenin, A.D.; Balakhontceva, M.A.; Ionov, M.V.; Kovalchuk, S.V. Extending Process Discovery with Model Complexity Optimization and Cyclic States Identification: Application to Healthcare Processes. *Algorithms* **2023**, *16*, 57. [CrossRef]
60. Mendling, J.; Reijers, H.A.; Cardoso, J. What makes process models understandable? In Proceedings of the Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, 24–28 September 2007; Proceedings 5; Springer: Heidelberg/Berlin, Germnay, 2007; pp. 48–63.
61. Reijers, H.A.; Mendling, J. A study into the factors that influence the understandability of business process models. *IEEE Trans. Syst. Man-Cybern.-Part Syst. Humans* **2010**, *41*, 449–462. [CrossRef]
62. Dikici, A.; Turetken, O.; Demirors, O. Factors influencing the understandability of process models: A systematic literature review. *Inf. Softw. Technol.* **2018**, *93*, 112–129. [CrossRef]
63. Hornix, P.T. Performance Analysis of Business Processes through Process Mining. Master's Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2007.
64. Laue, R.; Gadatsch, A. Measuring the understandability of business process models-Are we asking the right questions? In Proceedings of the Business Process Management Workshops: BPM 2010 International Workshops and Education Track, Hoboken, NJ, USA, 13–15 September 2010; Revised Selected Papers 8; Springer: Heidelberg/Berlin, Germany, 2011; pp. 37–48.
65. Imran, M.; Ismail, M.A.; Hamid, S.; Nasir, M.H.N.M. Complex process modeling in Process mining: A systematic review. *IEEE Access* **2022**, *10*, 101515–101536. [CrossRef]
66. Dzepina, A.; Lehner, F. Business Process Model Quality–Results from A Structured Literature Review. In Proceedings of the Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK, 23–28 June 2018.
67. de Oca, I.M.M.; Snoeck, M.; Reijers, H.A.; Rodríguez-Morffi, A. A systematic literature review of studies on business process modeling quality. *Inf. Softw. Technol.* **2015**, *58*, 187–205. [CrossRef]
68. Thilakaratne, M.; Falkner, K.; Atapattu, T. A systematic review on literature-based discovery workflow. *PeerJ Comput. Sci.* **2019**, *5*, e235. [CrossRef] [PubMed]
69. Polančič, G.; Cegnar, B. Complexity metrics for process models–A systematic literature review. *Comput. Stand. Interfaces* **2017**, *51*, 104–117. [CrossRef]
70. Sánchez González, L.; García Rubio, F.; Ruiz González, F.; Piattini Velthuis, M. Measurement in business processes: A systematic review. *Bus. Process. Manag. J.* **2010**, *16*, 114–134. [CrossRef]
71. Van Der Aalst, W.; Adriansyah, A.; De Medeiros, A.K.A.; Arcieri, F.; Baier, T.; Blickle, T.; Bose, J.C.; Van Den Brand, P.; Brandtjen, R.; Buijs, J.; et al. Process mining manifesto. In Proceedings of the Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, 29 August 2011; Revised Selected Papers, Part I 9; Springer: Heidelberg/Berlin, Germany, 2012; pp. 169–194.
72. Myung, I.J. The importance of complexity in model selection. *J. Math. Psychol.* **2000**, *44*, 190–204. [CrossRef]
73. Agrawal, R.; Gunopulos, D.; Leymann, F. Mining process models from workflow logs. In Proceedings of the Advances in Database Technology—EDBT'98: 6th International Conference on Extending Database Technology, Valencia, Spain, 23–27 March 1998; Proceedings 6; Springer: Heidelberg/Berlin, Germany, 1998; pp. 467–483.
74. Ferreira, D.R. *A Primer on Process Mining: Practical Skills with Python and Graphviz*; Springer: Cham, Switzerland, 2017.
75. Kummer, T.F.; Recker, J.; Mendling, J. Enhancing understandability of process models through cultural-dependent color adjustments. *Decis. Support Syst.* **2016**, *87*, 1–12. [CrossRef]
76. Blum, F.R. *Metrics in Process Discovery*; Technical Report TR/DCC-2015–6; Computer Science Department, University of Chile: Santiago, Chile, 2015.
77. Lassen, K.B.; van der Aalst, W.M. Complexity metrics for workflow nets. *Inf. Softw. Technol.* **2009**, *51*, 610–626. [CrossRef]
78. Cardoso, J. Complexity analysis of BPEL web processes. *Softw. Process. Improv. Pract.* **2007**, *12*, 35–49. [CrossRef]
79. Vergidis, K.; Tiwari, A.; Majeed, B. Business process analysis and optimization: Beyond reengineering. *IEEE Trans. Syst. Man, Cybern. Part (Appl. Rev.)* **2007**, *38*, 69–82. [CrossRef]
80. Tiwari, A.; Turner, C.; Majeed, B. A review of business process mining: State-of-the-art and future trends. *Bus. Process. Manag. J.* **2008**, *14*, 5–22. [CrossRef]
81. Wang, W.; Ding, H.; Dong, J.; Ren, C. A comparison of business process modeling methods. In Proceedings of the 2006 IEEE International Conference on Service Operations and Logistics, and Informatics, Shanghai, China, 21–23 June 2006; pp. 1136–1141.
82. Aguilar-Savén, R.S. Business process modelling: Review and framework. *Int. J. Prod. Econ.* **2004**, *90*, 129–149. [CrossRef]
83. Cardoso, J.; Mendling, J.; Neumann, G.; Reijers, H.A. A discourse on complexity of process models. In Proceedings of the Business Process Management Workshops, Berlin, Germany, 5–7 September 2006; Volume 4103, pp. 117–128. [CrossRef]
84. Mendling, J. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*; Springer Science & Business Media: New York, NY, USA, 2008; Volume 6.
85. Buijs, J.C.; van Dongen, B.F.; van der Aalst, W.M. Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *Int. J. Coop. Inf. Syst.* **2014**, *23*, 1440001. [CrossRef]
86. Buijs, J.C.; Van Dongen, B.F.; van Der Aalst, W.M. On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. In Proceedings of the OTM Conferences (1), Rome, Italy, 10–14 September 2012; Volume 7565, pp. 305–322.

87. Adriansyah, A.; Munoz-Gama, J.; Carmona, J.; van Dongen, B.F.; van der Aalst, W.M. Alignment based precision checking. In Proceedings of the Business Process Management Workshops: BPM 2012 International Workshops, Tallinn, Estonia, 3 September 2012; Revised Papers 10; Springer: Heidelberg/Berlin, Germany, 2013; pp. 137–149.
88. Adriansyah, A.; Munoz-Gama, J.; Carmona, J.; Van Dongen, B.F.; Van Der Aalst, W.M. Measuring precision of modeled behavior. *Inf. Syst.-Bus. Manag.* **2015**, *13*, 37–67. [CrossRef]
89. Mannhardt, F.; De Leoni, M.; Reijers, H.A.; Van Der Aalst, W.M. Measuring the precision of multi-perspective process models. In Proceedings of the Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, 31 August–3 September 2015; Revised Papers 13; Springer: Cham, Switzerland, 2016; pp. 113–125.
90. Remote Patient Monitoring Dataset-PMT Online. Available online: https://github.com/itmo-escience/ProFIT/blob/master/demo/log_examples/remote_monitoring_eng.csv (accessed on 30 April 2023).
91. van Dongen, B. Real-Life Event Logs-Hospital Log. 2011. Available online: https://data.4tu.nl/articles/_/12716513/1 (accessed on 30 April 2023).
92. Mannhardt, F. Hospital Billing-Event Log. 2017. Available online: https://data.4tu.nl/articles/_/12705113/1 (accessed on 30 April 2023).
93. Nurse Workflow in Hospitals-Almazov National Medical Research Centre. Available online: https://github.com/itmo-escience/ProFIT/blob/master/demo/log_examples/nurse_workflow.csv (accessed on 30 April 2023).
94. Mannhardt, F. Data-Driven Process Discovery-Artificial Event Log. 2016. Available online: https://data.4tu.nl/articles/_/12688325/1 (accessed on 30 April 2023).
95. Munoz-Gama, J.; de la Fuente, R.R.; Sepúlveda, M.M.; Fuentes, R.R. Conformance Checking Challenge 2019 (CCC19). 2019. Available online: https://data.4tu.nl/articles/_/12714932/1 (accessed on 30 April 2023).
96. Mannhardt, F. Sepsis Cases-Event Log. 2016. Available online: https://data.4tu.nl/articles/_/12707639/1 (accessed on 30 April 2023).
97. Tax, N.; Lu, X.; Sidorova, N.; Fahland, D.; van der Aalst, W.M. The imprecisions of precision measures in process mining. *Inf. Process. Lett.* **2018**, *135*, 1–8. [CrossRef]
98. Mendling, J.; Neumann, G.; Van Der Aalst, W. Understanding the occurrence of errors in process models based on metrics. *Lect. Notes Comput. Sci.* **2007**, *4803*, 113–130.
99. Munoz-Gama, J.; Carmona, J. Enhancing precision in process conformance: Stability, confidence and severity. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, Paris, France, 11–15 April 2011; pp. 184–191.
100. van Zelst, S.J.; Mannhardt, F.; de Leoni, M.; Koschmider, A. Event abstraction in process mining: Literature review and taxonomy. *Granul. Comput.* **2021**, *6*, 719–736. [CrossRef]
101. Rojas, E.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Process mining in healthcare: A literature review. *J. Biomed. Inform.* **2016**, *61*, 224–236. [CrossRef] [PubMed]
102. Munoz-Gama, J.; Martin, N.; Fernandez-Llatas, C.; Johnson, O.A.; Sepúlveda, M.; Helm, E.; Galvez-Yanjari, V.; Rojas, E.; Martinez-Millana, A.; Aloini, D.; et al. Process mining for healthcare: Characteristics and challenges. *J. Biomed. Inform.* **2022**, *127*, 103994. [CrossRef] [PubMed]
103. Richetti, P.H.P.; Baião, F.A.; Santoro, F.M. Declarative process mining: Reducing discovered models complexity by pre-processing event logs. In Proceedings of the Business Process Management: 12th International Conference, BPM 2014, Haifa, Israel, 7–11 September 2014; Proceedings 12; Springer: Cham, Switzerland, 2014; pp. 400–407.
104. van der Aalst, W.M. Process discovery: Capturing the invisible. *IEEE Comput. Intell. Mag.* **2010**, *5*, 28–41. [CrossRef]
105. McCabe, T. A Complexity Measure. *IEEE Trans. Softw. Eng.* **1976**, *SE-2*, 308–320. [CrossRef]
106. Cardoso, J. Evaluating the process control-flow complexity measure. In Proceedings of the IEEE International Conference on Web Services (ICWS'05), Orlando, FL, USA, 11–15 July 2005; p. 804. [CrossRef]
107. Cardoso, J. Control-flow complexity measurement of processes and Weyuker's properties. In Proceedings of the 6th International Enformatika Conference, Budapest, Hungary, 26–28 October 2005; Volume 8, pp. 213–218.
108. Henry, S.; Kafura, D. Software Structure Metrics Based on Information Flow. *IEEE Trans. Softw. Eng.* **1981**, *SE-7*, 510–518. [CrossRef]
109. Gruhn, V.; Laue, R. Adopting the cognitive complexity measure for business process models. In Proceedings of the 2006 5th IEEE International Conference on Cognitive Informatics, Beijing, China, 17–19 July 2006; Volume 1, pp. 236–241.
110. Shao, J.; Wang, Y. A new measure of software complexity based on cognitive weights. *Can. J. Electr. Comput. Eng.* **2003**, *28*, 69–74. [CrossRef]
111. Cardoso, J.; Vanderfeesten, I.; Reijers, H.A. Computing Coupling for Business Process Models. 2012. Available online: https://www.academia.edu/2895186/Computing_coupling_for_business_process_models (accessed on 30 April 2023).
112. Reijers, H.A.; Vanderfeesten, I.T. Cohesion and coupling metrics for workflow process design. In Proceedings of the Business Process Management: Second International Conference, BPM 2004, Potsdam, Germany, 17–18 June 2004; Proceedings 2; Springer: Heidelberg/Berlin, Germany, 2004; pp. 290–305.
113. Vanderfeesten, I.; Reijers, H.A.; Mendling, J.; van der Aalst, W.M.; Cardoso, J. On a quest for good process models: The cross-connectivity metric. In Proceedings of the Advanced Information Systems Engineering: 20th International Conference, CAiSE 2008, Montpellier, France, 16–20 June 2008; Proceedings 20; Springer: Heidelberg/Berlin, Germany, 2008; pp. 480–494.

114. Mendling, J.; Neumann, G. Error Metrics for Business Process Models. In Proceedings of the CAiSE Forum, Trondheim, Norway, 11–15 June 2007; pp. 53–56.
115. Yang, H.T.; Ju, J.H.; Wong, Y.T.; Shmulevich, I.; Chiang, J.H. Literature-based discovery of new candidates for drug repurposing. *Briefings Bioinform.* **2017**, *18*, 488–497. [CrossRef] [PubMed]
116. Vlietstra, W.J.; Zielman, R.; van Dongen, R.M.; Schultes, E.A.; Wiesman, F.; Vos, R.; Van Mulligen, E.M.; Kors, J.A. Automated extraction of potential migraine biomarkers using a semantic graph. *J. Biomed. Inform.* **2017**, *71*, 178–189. [CrossRef]
117. Shang, N.; Xu, H.; Rindflesch, T.C.; Cohen, T. Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J. Biomed. Inform.* **2014**, *52*, 293–310. [CrossRef]
118. Yetisgen-Yildiz, M.; Pratt, W. Using statistical and knowledge-based approaches for literature-based discovery. *J. Biomed. Inform.* **2006**, *39*, 600–611. [CrossRef] [PubMed]
119. Yetisgen-Yildiz, M.; Pratt, W. A new evaluation methodology for literature-based discovery systems. *J. Biomed. Inform.* **2009**, *42*, 633–643. [CrossRef] [PubMed]
120. Sánchez-González, L.; García, F.; Ruiz, F.; Mendling, J. A Study of the Effectiveness of Two Threshold Definition Techniques. In Proceedings of the 16th International Conference on Evaluation and Assessment in Software Engineering (EASE 2012), IET Seminar Digest, Ciudad Real, Spain, 14–15 May 2012.
121. Vanderfeesten, I.; Cardoso, J.; Mendling, J.; Reijers, H.A.; Van der Aalst, W. Quality metrics for business process models. *BPM Work. Handb.* **2007**, *144*, 179–190.
122. Laue, R.; Gruhn, V. Complexity Metrics for business Process Models. In Proceedings of the Business Information Systems, 9th International Conference on Business Information Systems, BIS 2006, Klagenfurt, Austria, 31 May–2 June 2006; Abramowicz, W., Mayr, H.C., Eds.; Volume P-85, pp. 1–12.
123. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the Ijcai, Montreal, Canada, 20 August 1995; Volume 14, pp. 1137–1145.
124. Van der Aalst, W.M. Verification of workflow nets. In Proceedings of the ICATPN, 23 June 1997; Volume 97, pp. 407–426.
125. Wang, N.; Sun, S.; OuYang, D. Business process modeling abstraction based on semi-supervised clustering analysis. *Bus. Inf. Syst. Eng.* **2018**, *60*, 525–542. [CrossRef]
126. Wang, Y.; Shao, J. Measurement of the cognitive functional complexity of software. In Proceedings of the The Second IEEE International Conference on Cognitive Informatics, London, UK, 20 August 2003; pp. 67–74.
127. Sánchez-González, L.; García, F.; Mendling, J.; Ruiz, F. Quality assessment of business process models based on thresholds. In Proceedings of the On the Move to Meaningful Internet Systems: OTM 2010: Confederated International Conferences: CoopIS, IS, DOA and ODBASE, Hersonissos, Crete, Greece, 25–29 October 2010; Proceedings, Part I; Springer: Heidelberg/Berlin, Germany, 2010; pp. 78–95.
128. Rolón, E.; Ruiz, F.; García, F.; Piattini, M. Applying software metrics to evaluate business process models. *CLEI-Electron. J.* **2006**, *9*, 5. [CrossRef]
129. Rolón, E.; García, F.; Ruiz, F.; Piattini, M.; Visaggio, C.A.; Canfora, G. Evaluation of BPMN models quality-a family of experiments. In Proceedings of the International Conference on Evaluation of Novel Approaches to Software Engineering, SCITEPRESS, Funchal, Madeira, Portugal, 4–7 May 2008; Volume 2, pp. 56–63.
130. Paakki, J.; Karhinen, A.; Gustafsson, J.; Nenonen, L.; Verkamo, A.I. Software metrics by architectural pattern mining. In Proceedings of the International Conference on Software: Theory and Practice (16th IFIP World Computer Congress), Beijing, China, 5–9 August 2000; pp. 325–332.
131. Woodward, M.R.; Hennell, M.A.; Hedley, D. A measure of control flow complexity in program text. *IEEE Trans. Softw. Eng.* 1979, 45–50. [CrossRef]
132. Cardoso, J. Business process quality metrics: Log-based complexity of workflow patterns. In Proceedings of the on the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS: OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS 2007, Vilamoura, Portugal, 25–30 November 2007; Proceedings, Part I; Springer: Heidelberg/Berlin, Germany, 2007; pp. 427–434.
133. Debnath, N.; Salgado, C.; Peralta, M.; Riesco, D.; Montejano, G. Optimization of the Business Process metrics definition according to the BPDM standard and its formal definition in OCL. In Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010, Hammamet, Tunisia, 16–19 May 2010; pp. 1–8.
134. Kluza, K.; Nalepa, G.J. Proposal of square metrics for measuring business process model complexity. In Proceedings of the 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), Wroclaw, Poland, 9–12 September 2012; pp. 919–922.
135. Cheng, C.Y. *Complexity and Usability Models for Business Process Analysis*; The Pennsylvania State University: State College, PA, USA, 2008.
136. Ghani, A.A.A.; Wei, K.; Muketha, G.M.; Wen, W.P. *Complexity Metrics for Measuring the Understandability and Maintainability of Business Process Models Using Goal-Question-metric (GQM)*; International Journal of Computer Science and Network Security; May 2008. Available online: http://paper.ijcsns.org/07_book/200805/20080533.pdf (accessed on 30 April 2023).
137. Syring, A.F.; Tax, N.; van der Aalst, W.M. Evaluating conformance measures in process mining using conformance propositions. *TTransactions on Petri Nets and Other Models of Concurrency XIV*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 11790, pp. 192–221.

138. Van der Aalst, W.; Adriansyah, A.; Van Dongen, B. Replaying history on process models for conformance checking and performance analysis. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 182–192. [CrossRef]
139. Goedertier, S.; Martens, D.; Vanthienen, J.; Baesens, B. Robust process discovery with artificial negative events. *J. Mach. Learn. Res.* **2009**, *10*, 1305–1340.
140. Leemans, S.J.; Fahland, D.; Van der Aalst, W.M. Scalable process discovery and conformance checking. *Softw. Syst. Model.* **2018**, *17*, 599–631. [CrossRef] [PubMed]
141. Weijters, A.; van Der Aalst, W.M.; De Medeiros, A.A. Process mining with the heuristics miner-algorithm. *Tech. Univ. Eindh. Tech. Rep. WP* **2006**, *166*, 1–34.
142. Polyvyanyy, A.; Solti, A.; Weidlich, M.; Di Ciccio, C.; Mendling, J. *Behavioural Quotients for Precision and Recall in Process Mining*; Technical Report; University of Melbourne: Melbourne, Australia, 2018.
143. Rozinat, A.; Van der Aalst, W.M. Conformance checking of processes based on monitoring real behavior. *Inf. Syst.* **2008**, *33*, 64–95. [CrossRef]
144. Munoz-Gama, J.; Carmona, J. A fresh look at precision in process conformance. In Proceedings of the Business Process Management: 8th International Conference, BPM 2010, Hoboken, NJ, USA, 13–16 September 2010; Proceedings 8; Springer: Heidelberg/Berlin, Germany, 2010; pp. 211–226.
145. vanden Broucke, S.K.; De Weerdt, J.; Vanthienen, J.; Baesens, B. Determining process model precision and generalization with weighted artificial negative events. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1877–1889. [CrossRef]
146. van Dongen, B.F.; Carmona, J.; Chatain, T. A unified approach for measuring precision and generalization based on anti-alignments. In Proceedings of the Business Process Management: 14th International Conference, BPM 2016, Rio de Janeiro, Brazil, 18–22 September 2016; Proceedings 14; Springer: Cham, Switzerland, 2016; pp. 39–56.
147. Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A. Methodologies for data quality assessment and improvement. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–52. [CrossRef]
148. Reijers, H.; Mendling, J. Modularity in process models: Review and effects. In Proceedings of the Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, 2–4 September 2008; Proceedings 6; Springer: Heidelberg/Berlin, Germany,2008; pp. 20–35.
149. Soo, L.G.; Jung-Mo, Y. An empirical study on the complexity metrics of petri nets. *Microelectron. Reliab.* **1992**, *32*, 323–329. [CrossRef]