2II66 Assignment 3 – Robbert Jongeling – 0747896

# //TODO discovery: Can you indicate why you believe that the model is a good description of the behavior captured in the log?

# Introducion:

### We investigate the processes before and after the visit of the *bucky room* modality.

We consider an event log from the Isala hospital. The log contains events for four different locations of the hospital. These are Vlinder (V), Diagnosis (D), OC Heerde (H) and OC Kampen (K). We investigate the processes in the hospitals before and after the *bucky room* modality. This event is concerned with taking x-ray pictures.

# Process Discovery

We discover the process for the given event log by filtering the log APM\_Third\_Assignment\_Before.xes for processes before visit of the modality and APM\_Third\_Assignment\_After.xes for processes after visit of the modality. We filter using the *Filter Log by Attributes* plugin in Prom6. We use the following settings: **filter on***: trace with an event having this attribute,* **attribute***: concept:name.* The **value** differs per location and is for location X: RADIO:Radio-X-bucky room. We **keep matching traces.**

The resulting discovered model when using the Inductive miner with noise level 0% yields the following model for location V before visit of the modality:



Figure : Too large model of the entire log for location V

It is clear that such a model is not understandable. We therefore filter the logs such that they only include the 15 most frequently occurring events. We do this by applying the *Filter Log using Simple Heuristics* plugin of Prom6. We select all start and end events and in the *event filter*, we choose a percentage such that we keep the 15 most frequently occurring events.

## Can you discover *per location* what is the process *before* visit of the modality?

### Location V:

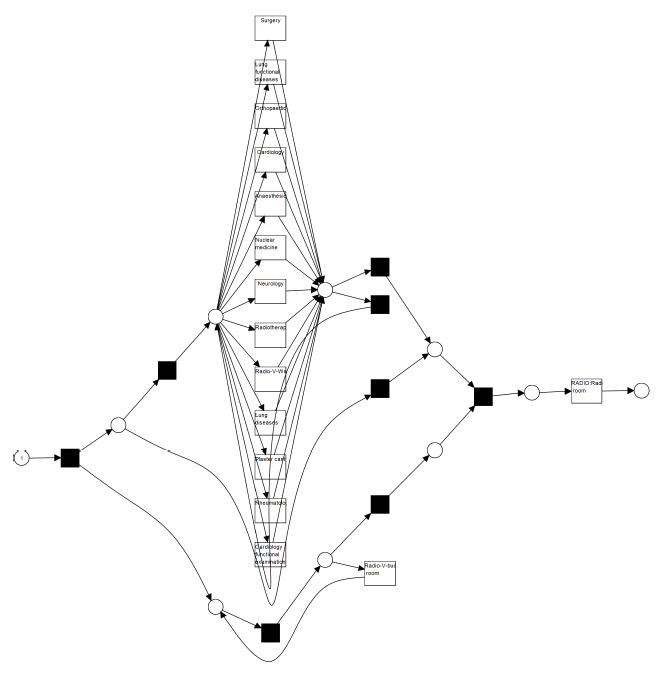
We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 1. The fitness of this model is 0.962. It is a clear model, it shows a bunch of other departments that can be visited zero or more times before visiting the bucky room. This is a good description of the behaviour captured in the log. The log shows, as also seen in the too large model in Figure 1, a number of different departments that are visited before the bucky room.

Figure : Discovered model for location V before visit of the modality

### Location D:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 2. The fitness of this model is 0.988. Its structure is similar to the process before visiting the modality at location V. And is also a good description of the behaviour in the log for the same reasons as mentioned there.

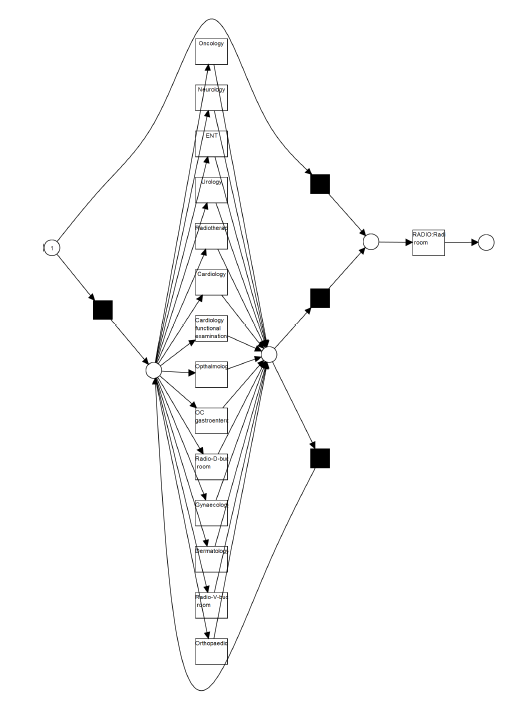


Figure : Discovered model for location D before visit of the modality

### Location H:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 3. The fitness is 1. The general structure is similar to the previous two discovered models and is also a good description of the behaviour in the log for the same reasons as mentioned there. The difference is that we now sometimes see an order in the events prior to the *bucky* *room* modality. For example, the *Psychology* event is always prefaced with a visit to *Neuroloty* or *Anaesthesia*.

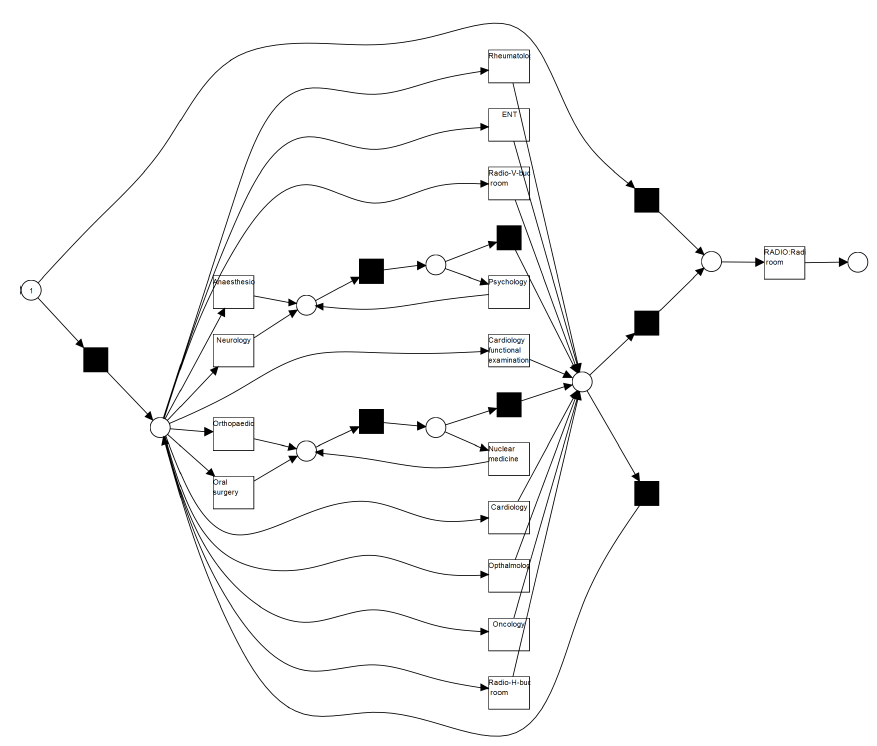


Figure : Discovered model for location H before visit of the modality

### Location K:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 4. The fitness of this model is 0.982. Its structure is similar to that of the models for locations V and D And is also a good description of the behaviour in the log for the same reasons as mentioned there.

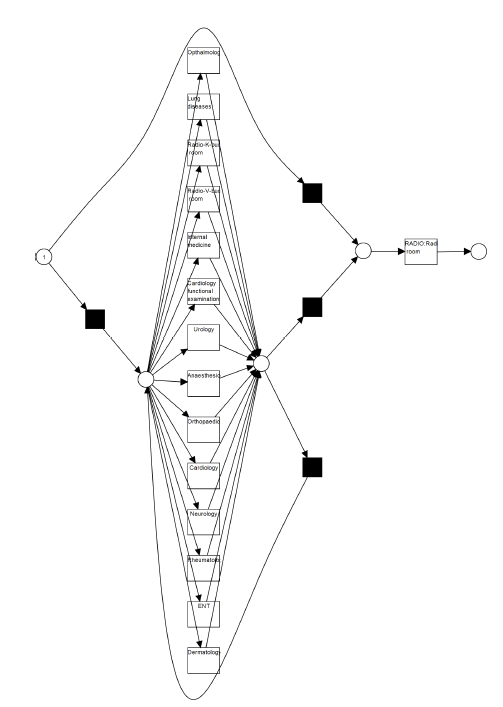


Figure : Discovered model for location K before visit of the modality

## Can you discover per location what is the process *after* visit of the modality?

### Location V:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 6. The model fitness is 0.971.

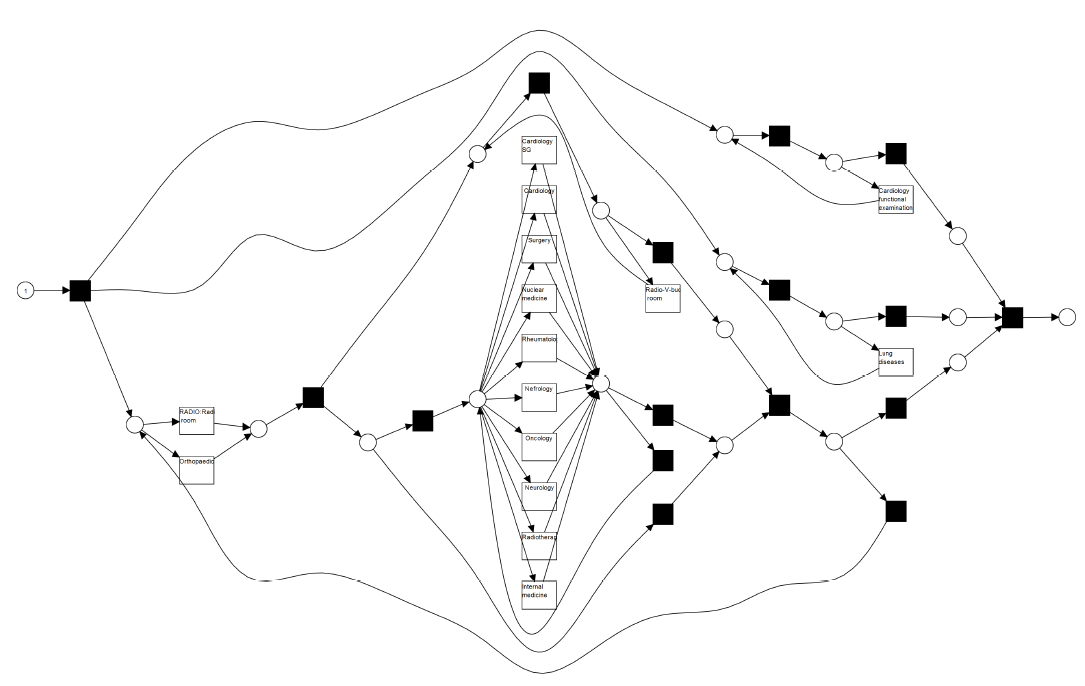


Figure : Discovered model for location V after visit of the modality

### Location D:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.1. The resulting model is shown in Figure 7. The fitness is 0.948.

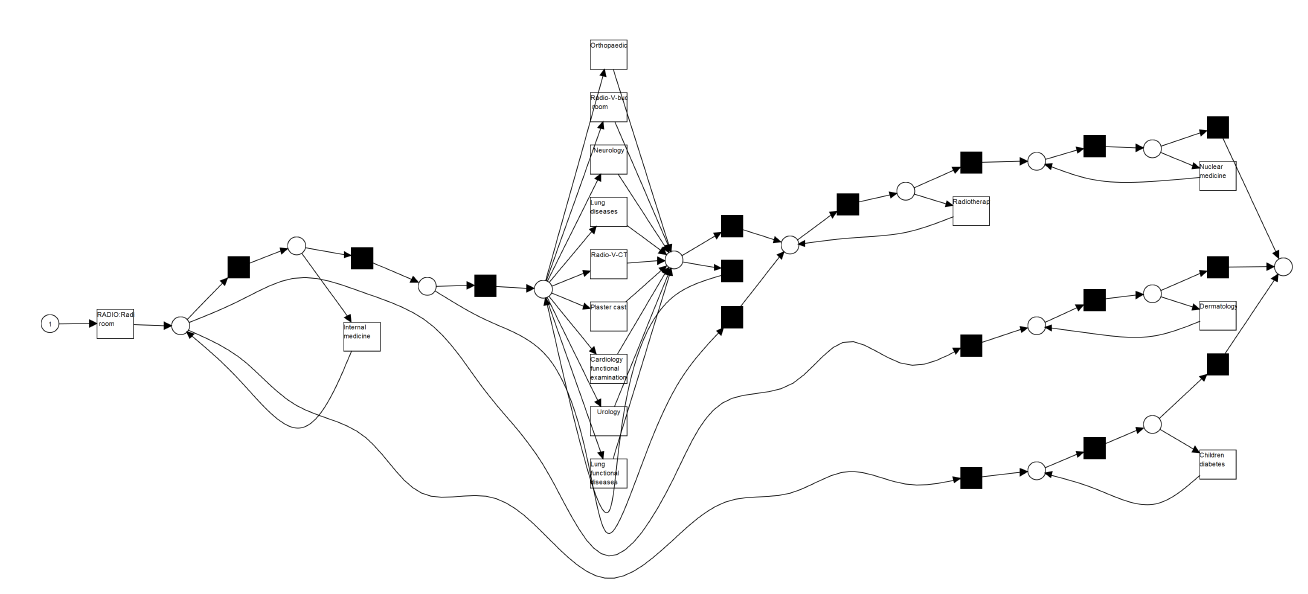


Figure : Discovered model for location D after visit of the modality

### Location H:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.2. The resulting model is shown in Figure 8. The fitness is 0.848.

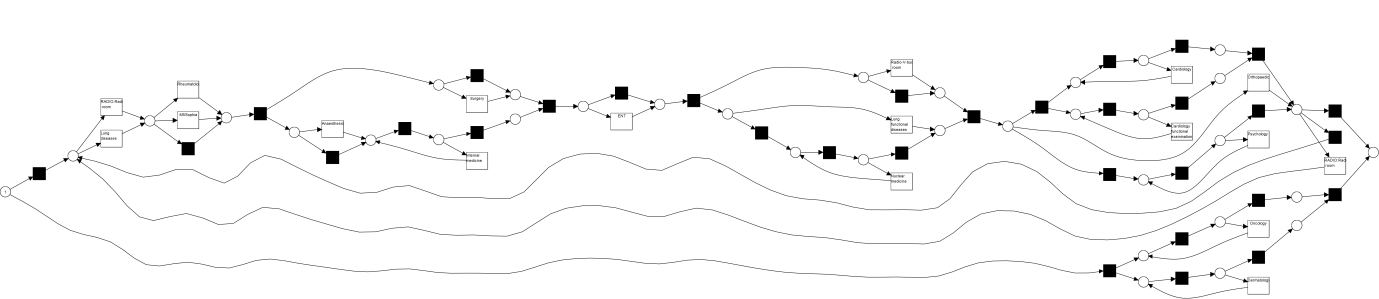


Figure : Discovered model for location H after visit of the modality

### Location K:

We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.1. The resulting model is shown in Figure 9. The fitness is 0.934.

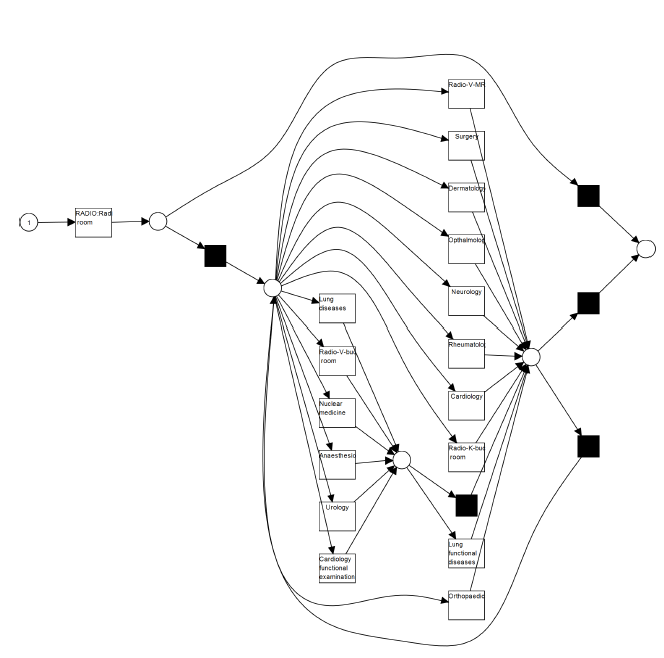


Figure : Discovered model for location K after visit of the modality

# Bottlenecks

For a task to be a bottleneck, it must satisfy two conditions. Its throughput time is high compared to the other tasks in the process. Second, it occurs frequently. A task that takes 10 hours in a process where all the others take 1 hour is only a bottleneck if it occurs frequently. For example, if the 10 hour task is a yearly reoccurring event and the other tasks are daily, then we do not consider it to be a bottleneck.

## Can you indicate *per location* what are the top 3 of bottlenecks within the process *before* visit of the modality?

We now look at the bottlenecks in the processes for every location before the visit of the modality.

### Location V:

### Location D:

### Location H:

### Location K:

## Can you indicate *per location* what are the top 3 of bottlenecks within the process *after* visit of the modality?

We look at the bottlenecks (as defined in Q3) in the processes for every location after the patient visits the modality.

### 14) Location V:

### 15) Location D:

### 16) Location H:

### 17) Location K:

# Model comparison

To find out whether some of the locations have comparable processes, we compare their models. We replay the (filtered) log of each location on the discovered model of each location. The obtained fitness values are shown in Table 1 and Table 2. For two models to be considered comparable, their pairwise fitness should be high. E.g. the fitness of log V on model D should by high as well as the fitness of log D on model V. On the diagonal of the tables, we see the logs replayed on their own models, so these are not surprisingly high.

## 18) *Before* the visit to the modality, can you indicate whether the processes of the 4 locations are comparable to each other?

The results of replaying the logs for each location on the discovered models of each location are shown in Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model (V) | Model (D) | Model (H) | Model (K) |
| Log (V) | 0.962 | 0.256 | 0.323 | 0.354 |
| Log (D) | 0.053 | 0.988 | 0.079 | 0.099 |
| Log (H) | 0.210 | 0.215 | 1 | 0.244 |
| Log (K) | 0.127 | 0.131 | 0.133 | 0.982 |

Table : fitness of logs when replayed on models for processes before visiting modality

The processes are not comparable to each other. Their pairwise fitness is not close to 1 or even higher than 0.4.

We distinguish two reasons for this low pairwise fitness. The first is that the processes at the different locations are not at all similar, this is not very likely. We would expect some degree of similarity between the processes even at different locations as we have seen such similarity in the structure of the discovered models.

The second possible reason is that by filtering the processes to have only the 15 most occurring events, we probably have created dissimilarity between the locations. Some locations may be specialized in certain events which makes them occur there much more frequent than anywhere else. Which would make it to the model for that location but not to the model for the other locations. This second possibility is much more likely.

## 19) *After* the visit to the modality, can you indicate whether the processes of the 4 locations are comparable to each other?

The results of replaying the logs for each location on the discovered models of each location are shown in Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model (V) | Model (D) | Model (H) | Model (K) |
| Log (V) | 0.971 | 0.289 | 0.167 | 0.326 |
| Log (D) | 0.173 | 0.948 | 0.089 | 0.169 |
| Log (H) | 0.274 | 0.204 | 0.848 | 0.273 |
| Log (K) | 0.193 | 0.166 | 0.078 | 0.934 |

Table : fitness of logs when replayed on models for processes after visiting modality

The processes are not comparable to each other. Their pairwise fitness is not close to 1 or even higher than 0.4. The reasons are the same as the reasons provided for the processes before visiting the modality (Q18).