2II66 Assignment 3 – Robbert Jongeling – 0747896

# //TODO discovery: Can you indicate why you believe that the model is a good description of the behaviour 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:

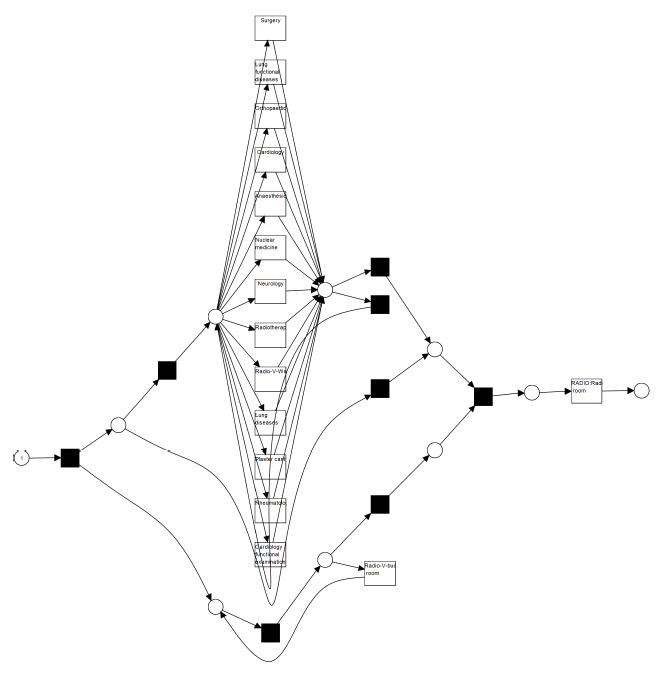
We mined the filtered log as described above using the Inductive miner – infrequent, with noise threshold 0.0. The resulting model is shown in Figure 1. The fitness of this model is 0.962. It is a clear model, it shows a bunch of other departments that can be visited zero or more times before visiting 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.

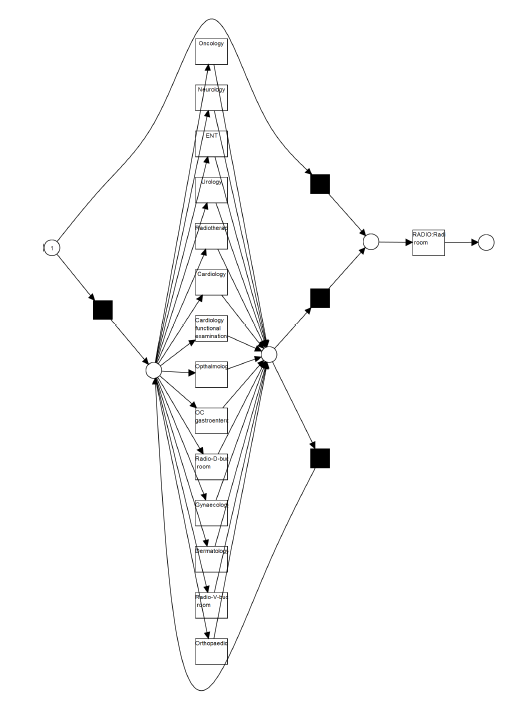


Figure : Discovered model for location D before visit of the modality

### 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*.

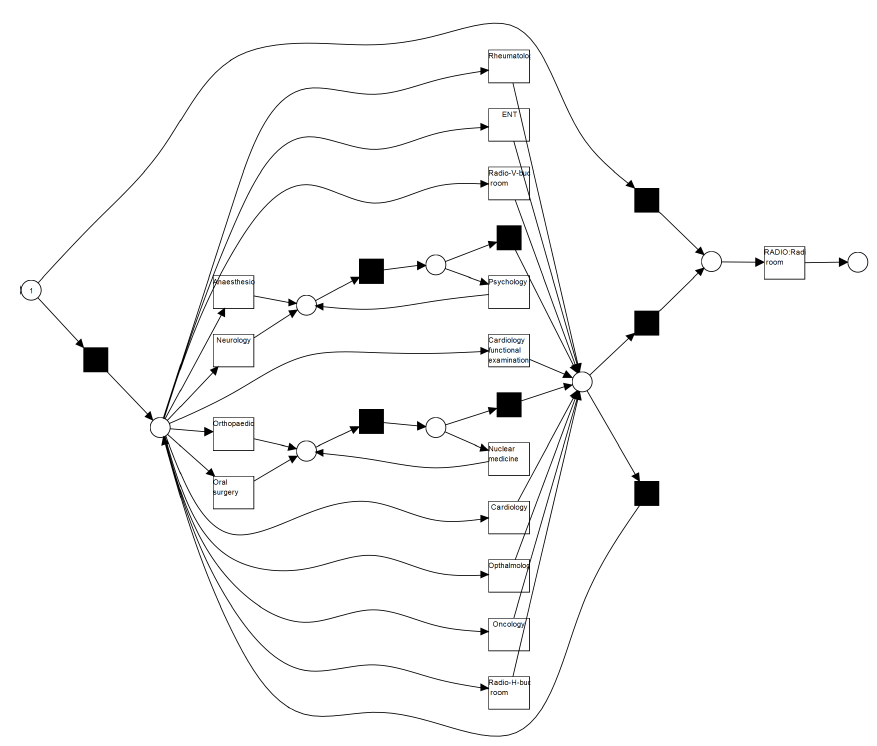


Figure : Discovered model for location H before visit of the modality

### 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.

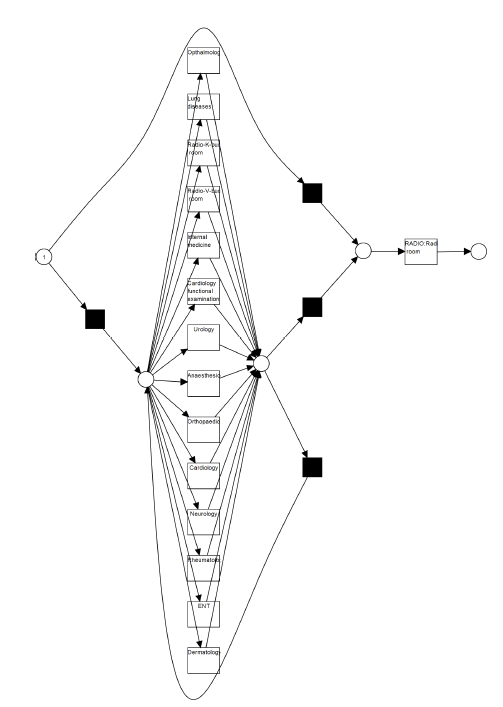


Figure : Discovered model for location K before visit of the modality

## 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.

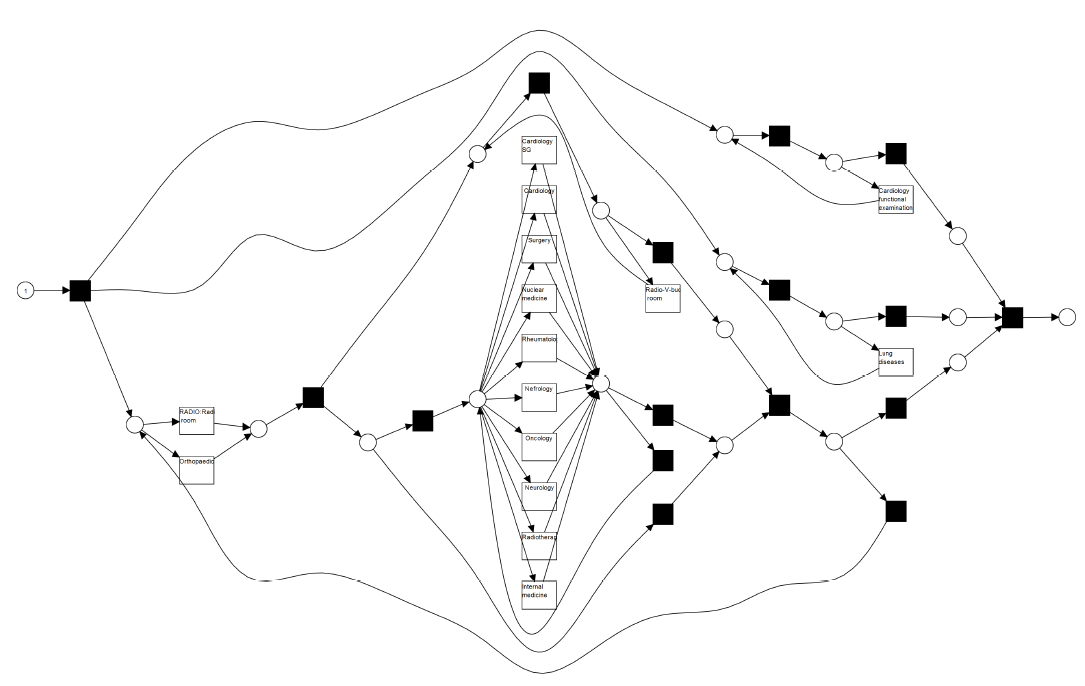


Figure : Discovered model for location V after visit of the modality

### 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.

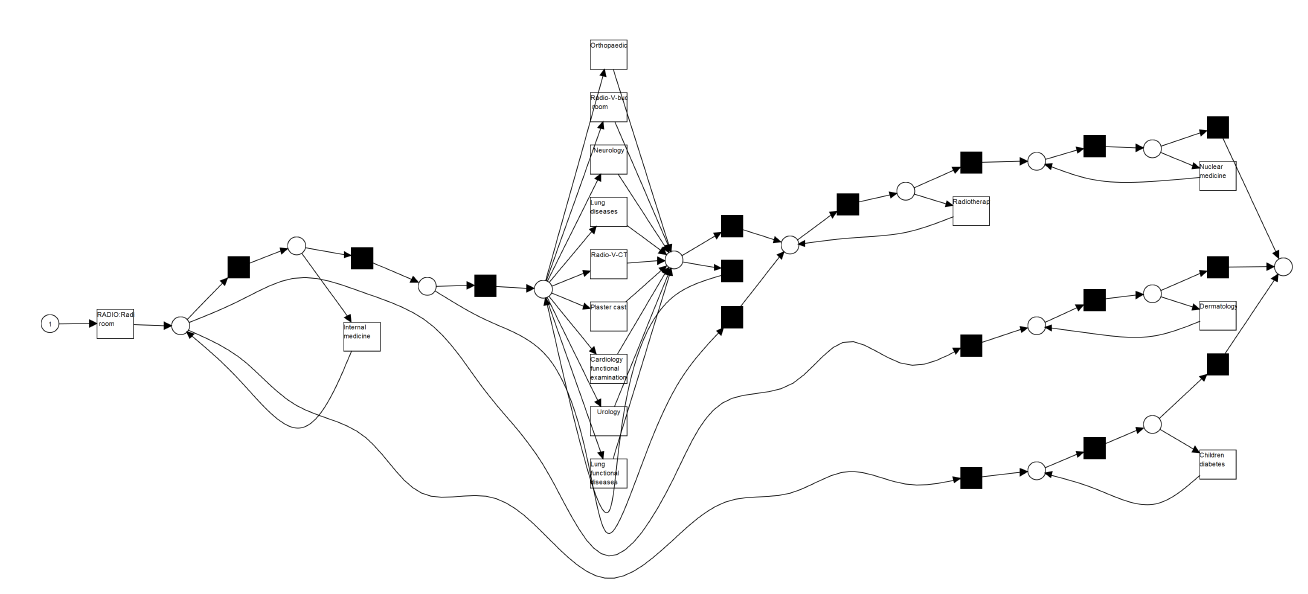


Figure : Discovered model for location D after visit of the modality

### 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.

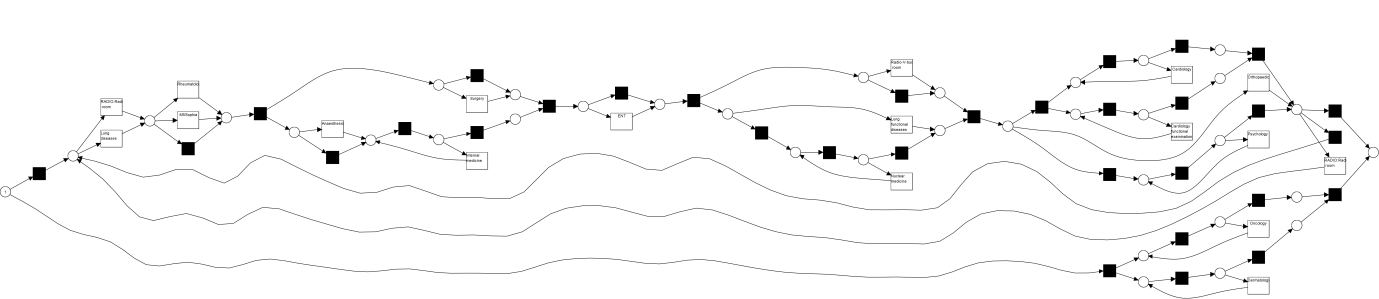


Figure : Discovered model for location H after visit of the modality

### 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.

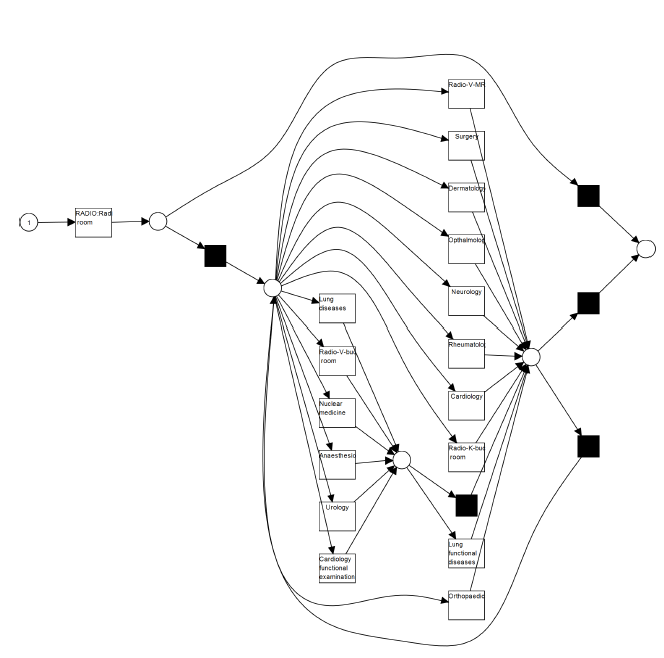


Figure : Discovered model for location K after visit of the modality

# Bottlenecks

For a task to be a bottleneck, it must satisfy two conditions. Its average waiting 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. We consider the models as discovered in the first part of this report. In this analysis we do not consider the visit to the bucky room itself as a bottleneck, because we are interested in the process before that visit.

### Location V:

Neurology, Orthopaedics and Surgery are the tree tasks with the highest average waiting time. Of these, Orthopaedics occurs frequently when compared to the other tasks. (334 times, avg. waiting time 13.34 days) The other two tasks occur a little over 100 times and are “beaten” by other tasks that have shorter avg. waiting time but occur more often. The Cardiology event occurs 474 times and has an avg. waiting time of 10.69 days. The Lung diseases event occurs 306 times and has an avg. waiting time of 10.09 days.

So the tree largest bottlenecks for location V in the process before visiting the bucky room modality are the Orthopaedics, Cardiology and Lung diseases events.

### Location D:

The Neurology and Gastroenterology events have the highest average waiting time (13.6 days) but occur not often (21 and 18 times respectively). These are not bottlenecks. The Radiotherapy event occurs the second most often, at 61 occurrences. But its low avg. waiting time (2.94 days) makes this event also not a bottleneck.

The Dermatology event has 62 occurrences and an avg. waiting time of 7.96 days. Dermatology is of these two the bigger bottleneck. The Cardiology event has 43 occurrences and has an avg. waiting time of 8.35 days. The third large bottleneck is the Urology event, with an avg. waiting time of 10.21 days and with 25 occurrences.

So the tree largest bottlenecks for location D in the process before visiting the bucky room modality are the Dermatology, Cardiology and Urology events.

### Location H:

For this location, there is a relatively low number of cases, 58. Orthopaedics is the event with the highest average waiting time (15.51 days) and occurs relatively often (16 times). Another event with a high average waiting time is Anaesthesiology (14.86 days) but this event only occurs 7 times so we do not consider it a bottleneck. The Oncology event has a much shorter average waiting time (10.34 days) but occurs more often (21 times). We do consider this event a bottleneck. The third bottleneck is the Cardiology functional examination event, with an avg. waiting time of 9.59 days and 12 occurrences.

So the tree largest bottlenecks for location H in the process before visiting the bucky room modality are the Orthopaedics, Oncology and Cardiology functional examination events.

### Location K:

As at location H, the frequency of the events is relatively low. There is little difference in number of occurrences per event and thus the avg. waiting time is the dominant factor when determining if an event is a bottleneck. The Orthopaedics event is the event with high avg. waiting time (19.54 days) and is the event with the highest number of occurrences (34). We naturally consider this a bottleneck. The Internal medicine event only occurs 12 times but has an avg. waiting time of 21.93 days and is thus also considered a bottleneck. The rheumatology event is the last bottleneck, with an average waiting time of 15.69 days (still relatively high for the events in this process) and 25 occurrences

So the tree largest bottlenecks for location K in the process before visiting the bucky room modality are the Orthopaedics, Internal medicine and rheumatology events.

## 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. We do not consider events before or in parallel with the bucky room event or the bucky room event itself, we are only interested in the bottlenecks strictly after it.

### Location V:

We see here again the same as we have noted before, the number of occurrences of the events is very similar and thus the average waiting time dominates the decision to call some events bottlenecks and other not. The event with the highest average waiting time is Surgery (1.10 months). Although it occurs relatively infrequent (142 times on 984 total cases) it is a bottleneck. Neurology also has a high avg. waiting time (28.99 days) and although this also has a low number of occurrences (118) we consider it a bottleneck. The last bottleneck is Cardiology, with 161 occurrences it is more frequent than the previous but its waiting time is less, 23.91 days.

So the tree largest bottlenecks for location V in the process after visiting the bucky room modality are the Orthopaedics, Internal medicine and rheumatology events.

### Location D:

Again, the frequency of the different events is similar and the average waiting time dominates the decision to declare some events bottlenecks. But not completely, the event Internal medicine has an avg. waiting time of 1.06 months but only occurs 22 times. The Neurology event has a much lower average waiting time of 21.71 days but occurs almost twice as much (41 times). We consider Neurology a bigger bottleneck than Internal medicine. Other bottlenecks are Orthopaedics (occurs 73 times, avg. waiting time is 25.92 days) and Cardiology functional examination (occurs 51 times, avg. waiting time is 28.84 days).

So the tree largest bottlenecks for location D in the process after visiting the bucky room modality are the Neurology, Orthopaedics and Cardiology functional examination events.

### 18) Location H:

Considering the average waiting time, Cardiology functional examination is the biggest bottleneck with an avg. waiting time of 1.88 months. But as there are only 3 occurrences of it, we do not consider it a bottleneck. An event with a lower but still relatively high avg. waiting time is Radio-V-bucky room (27.20 days). It occurs 20 times which is not a lot but none of the events occur a lot of times. We consider this a bottleneck. At an even lower number of cases (11) is Orthopaedics, but with its avg. waiting time of 1.11 months, it is a bottleneck. The last bottleneck is Cardiology, with 14 occurrences and 26.36 days avg. waiting time.

So the tree largest bottlenecks for location H in the process after visiting the bucky room modality are the Radio-V-bucky room, Orthopaedics and Cardiology events.

### Location K:

At this location, the average waiting time of most events is high (around 20 days). So here, the frequency dominates. Orthopaedics has 28.65 days waiting time (the longest) and is also the most frequent event at 46 occurrences. Naturally, we consider this a bottleneck. Rheumatology has an avg. waiting time of 21.42 which is a bit lower than most but its frequency is 31 times, which makes it a larger bottleneck than others. The last bottleneck we name is Cardiology, with a high avg. waiting time of 28.30 days and not that high frequency of 20.

So the tree largest bottlenecks for location K in the process after visiting the bucky room modality are the Orthopaedics, Rheumatology and Cardiology events.

# 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).