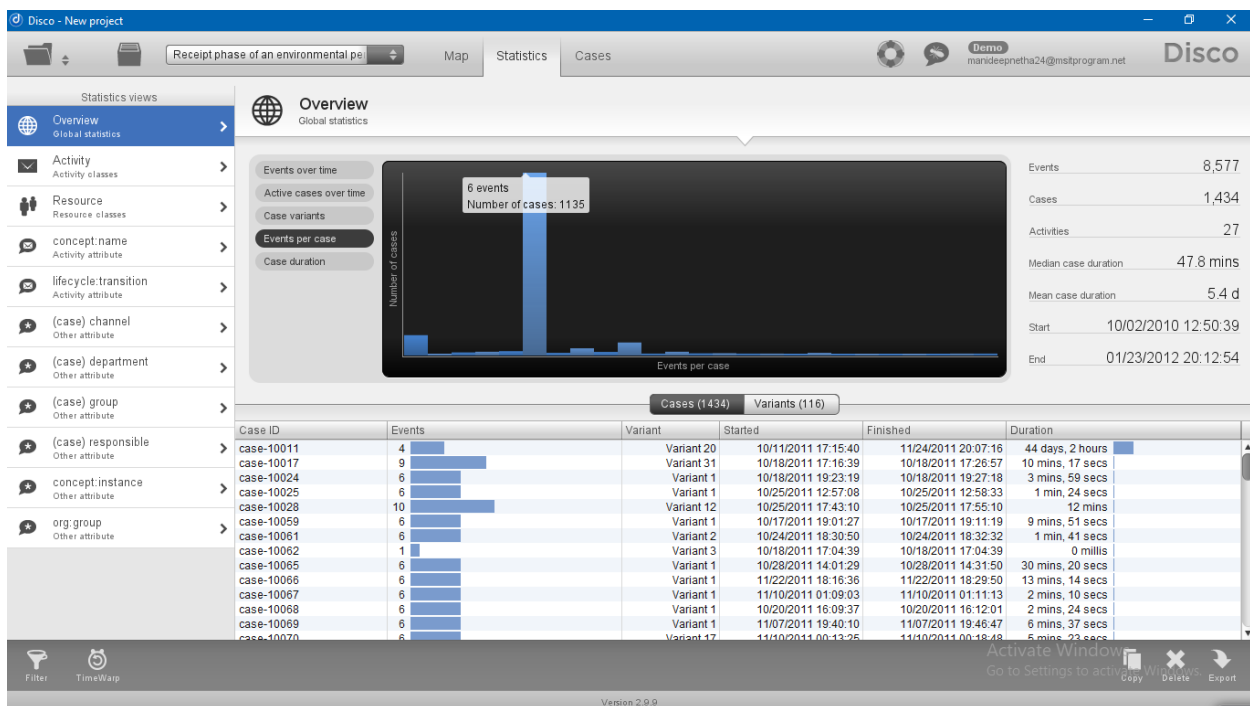


1. Average events per case

Approach I used: After loading the data into Disco I switched to the "Statistics" tab and clicked on "Events per case" to the left of the graph to show the relevant graph in order to verify my calculation of the average events per case with the graphical representation.

What I saw: From the summary statistics to the right | divided the number of Events by the number of Cases, i.e., $8'577/1434$ resulting in 5.98 average events per case. The mode, the value that occurred most often, calculated mean. See screenshot shown in the graph is 6, corresponding to the

My analysis: Based on this the average number of events per case is 5.98.



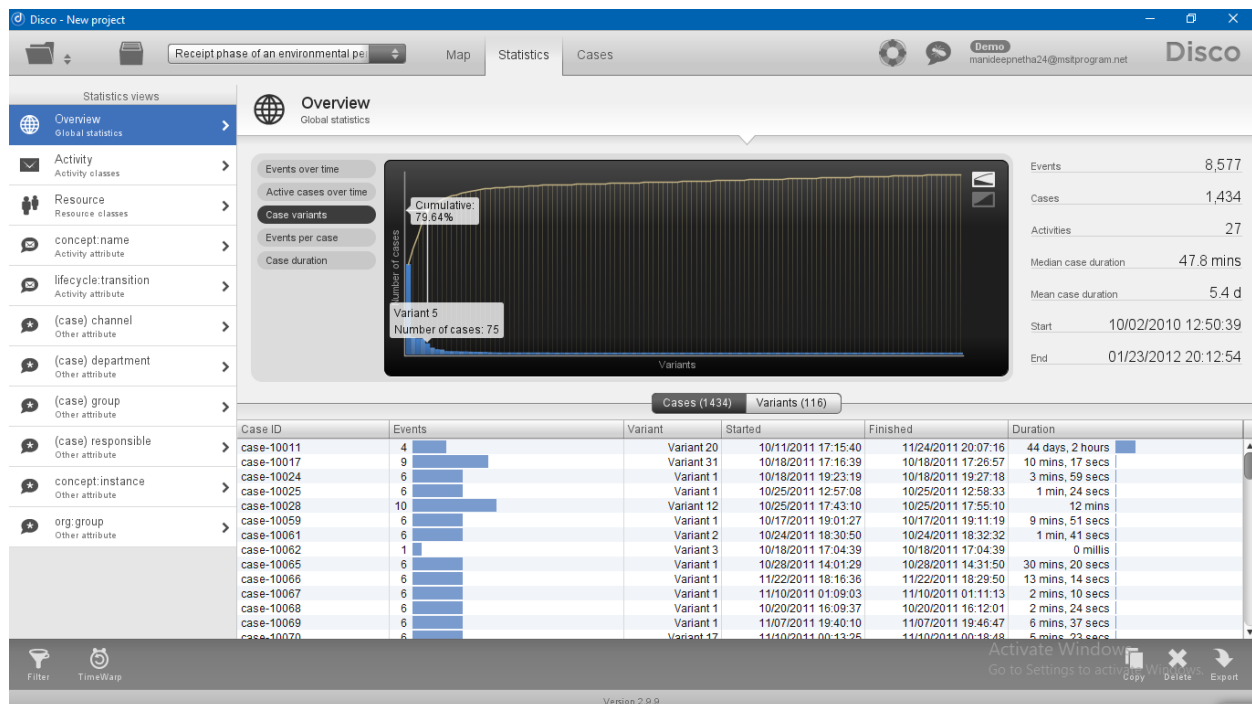
2. Unique cases or many cases following the same activity sequence

Approach I used: While in the "Statistics" tab I switched to "Case variants" to the left of the graph in order to analyse whether each case seems to be unique or whether many cases follow the same activity sequence.

What I saw: The graph showed that out of a total of 116 variants one case variant with a specific activity sequence covering 50% of the cases, i.e., 713 divided by the total number of cases 1'434. See screenshot

Moving the cursor along the cumulative percentage curve on the upper side of the curve shows that the first five case variants to the left with five different activity sequence variants cover around 80% of the cases. See screenshot

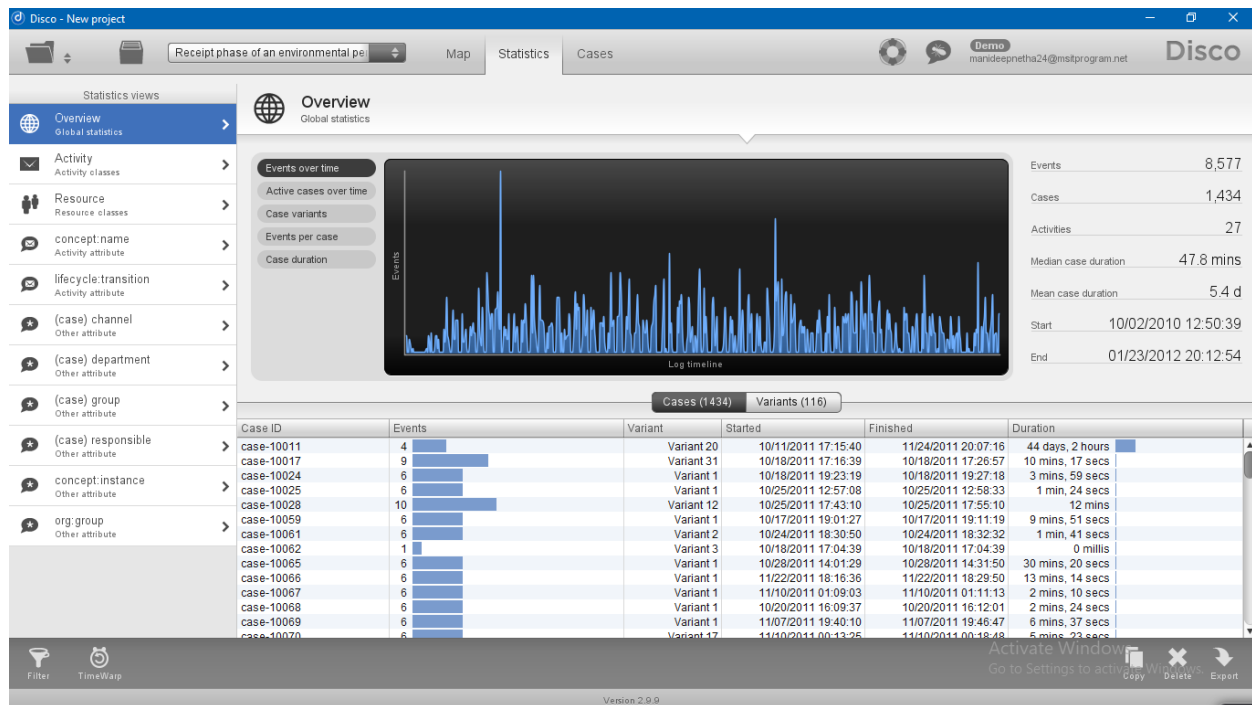
My analysis: Many cases are not unique. Half the cases follow one activity sequence, 80% of the cases are covered by five of a total of 116 case variants.



3. Main observation from the 'Events over time' graph

Approach I used: While still in the "Statistics" tab I switched to "Events over time" to the left of the graph and then scrolled alongside the x-axis for analyzing observations. As I did not find any main observation at the beginning I exported the event log as CSV file, showed the data in separate columns and formatted the timestamp such that it included weekdays. What I saw: I saw in the graph that the events are not even over time. Sometimes there are spikes at the end of the month or in the morning hours. With the help of the event log export I saw that most of the time there is no events on the weekends.

My analysis: The main observation is that there is almost no event on weekends.



2.

Approach I used: After loading the data and completing the exercise on Statistics in Disco, I

switched to the Map view and put the activity detail slider to 0% (the lowest) in order to focus on the most frequent from the 27 activities and the paths slider to 20% (rather low). This shows the main paths between the most frequent activities. I chose this approach as there was 116 activity sequence variants of the 27 events, which resulted in a very complex graph which would have been difficult to interpret.

What I saw: I saw a process map with 6 activities and 15 arcs connecting them. The process map was clearly readable, the common activities were darker blue, and the more frequent paths had thicker arrows. I furthermore saw that there are various loops back 3 from activity T10 Determine necessity to stop indication - complete'. 2 from activity 'T06 Determine necessity of stop advice - complete' and one from activity T04 Determine confirmation of receipt complete'. Activity T02 Check confirmation of receipt - complete shows a loop in itself indicating that the activity is repeated in 51 cases.

My analysis: Based on the process map, can conclude that the main path consists of the sequence Determine necessity of stop advice - complete' -> 'T10

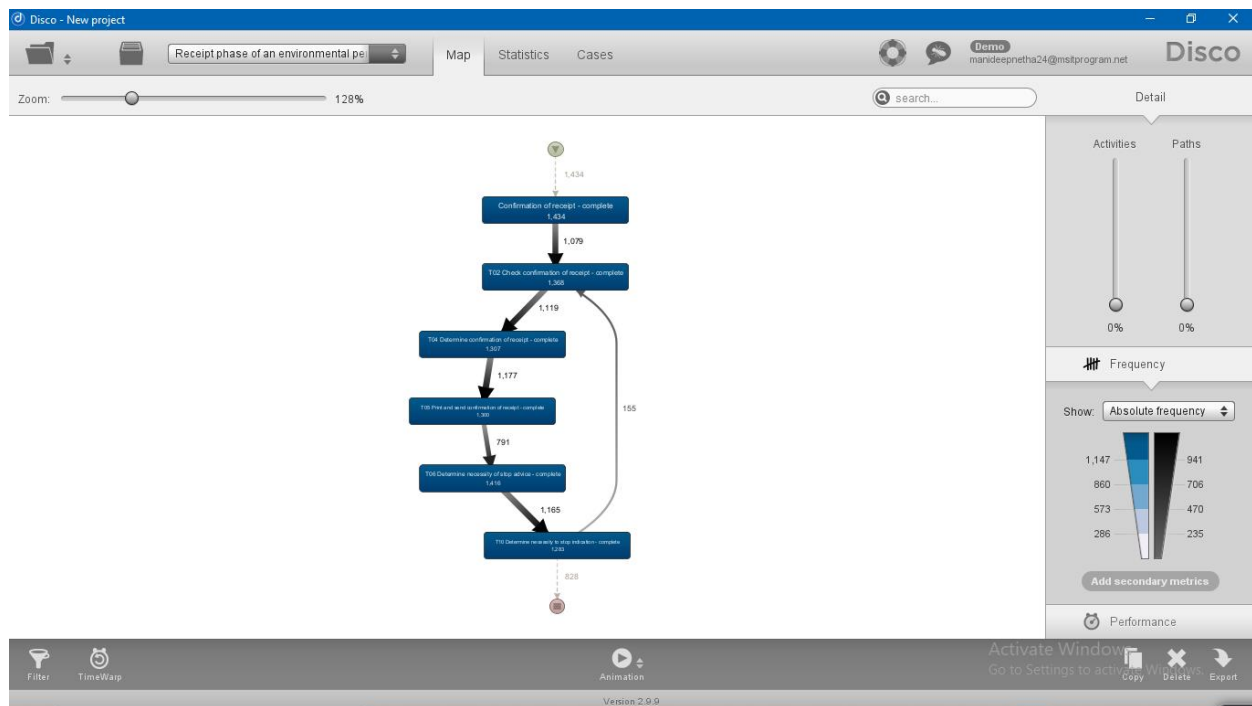
'Confirmation of receipt - complete-> T02 Check

confirmation of receipt complete > T04 Determine

confirmation of receipt - complete-> T05 Print and

send confirmation of receipt - complete 'T06

Determine necessity to stop indication - complete.



2. Frequent activities and paths between activities

Approach I used: I kept the settings activity slider at 0% (the lowest) in order to focus on the most frequent activities and the path slider at 20% (rather low). What I saw: I saw a process map with the 6 most frequent activities identical to the ones from question one with respect to the main process. With zooming in on the most frequent path I identified the paths of the main process flow described in question 1 as the most frequent ones. See screenshot

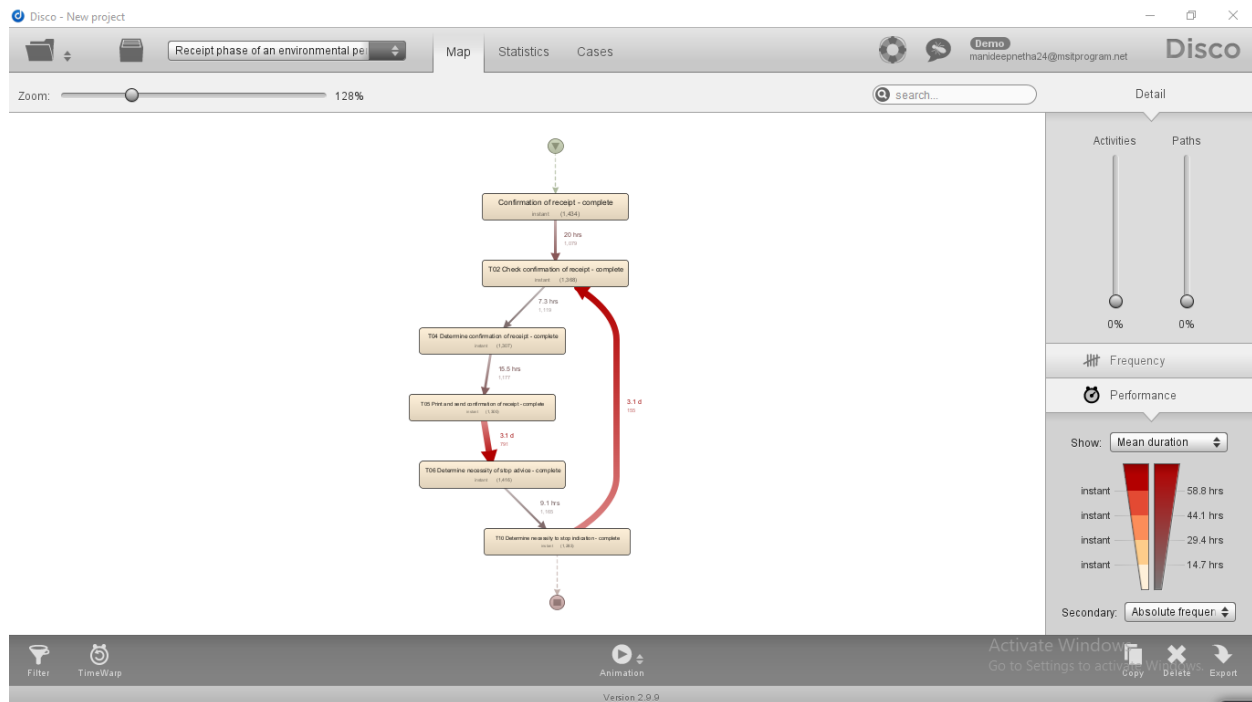
My analysis: Based on the process map and the respective settings and analysis described above I can conclude that the main activities are 'Confirmation of receipt - complete, T02', 'T04 T05, T06' and T10. The main paths are the ones from the main process, i.e.

Start -> Confirmation of receipt complete -> 'T02 -> 'T04 ->'T05 ->'T06 ->'T10

Performance Projection I kept the activity slider on 0% (the lowest) and the path slider on 20% (rather low). I switched from "Frequency" to "Performance" with the clock-symbol at the right side of the screen. In the bottom right corner I first selected "Mean duration" instead of "Total duration" next to "Show". I subsequently switched to "Max. duration" in order to analyse the process steps with long times.

What I saw: The longest mean duration was the loop in itself for activity T02' with 5.8 d. There were three process steps with mean durations of 3.1 d, T05 to 'TOG' from the main process and the loops back from T10' to T02 as well as from T04 to T02. The other process steps of the main flow were all below 3 d. See screenshot

In order to drill down on the long process steps, I saw that the max. duration for TOS to T06' was 38.4 w and the one for the loop back from T04' to T02 15.3 w, whereas the max. duration for the loop back from T10 to 'T02 is only 71.2 d. i.e., around 10 w.



My analysis: Combining the findings from analyzing mean and max, duration I suggest to analyze the 155 cases with loops back from T10 to T02 and T04 to T02' as they take very long with all measures. For the process step 'T05 to T06' the "long runners" should be analyzed with respect to root cause analysis.

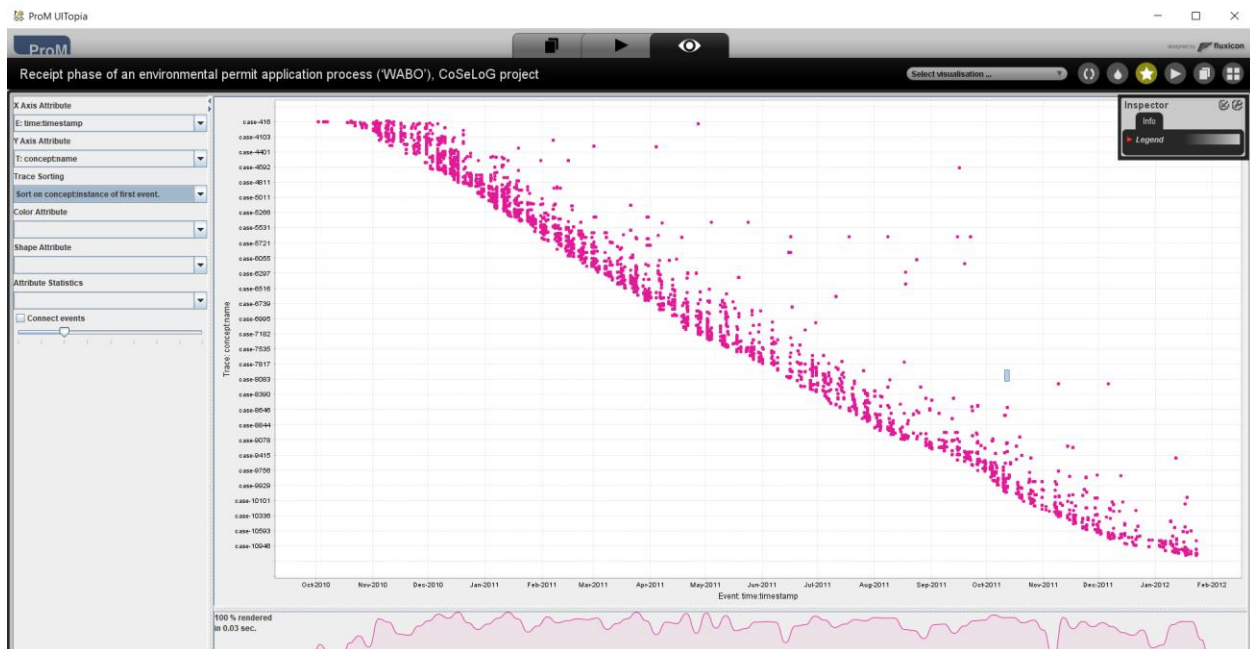
1. Constant arrival rate

Approach I used: After importing the XES file into Prom I selected the "eye symbol" with the event log selected in the Workspace - Tab "All" in the upper right corner for function "View resource". I then selected "Dotted Chart (Log Projection, under "Select visualization" in the upper right-hand corner of the screen, changed sorting to "Sort on time: timestamp of first event" and Color Attribute to "C: Activity classifier.

What I saw I saw a dotted chart with time on the x- axis and cases on the y-axis, whereby there was a

continuous straight line from the top left to the bottom right of the graph indicating a constant arrival rate most of the time. There is only a slight reduce in the arrival rate at the end of the timeline.

My analysis: As the line from the upper left to the lower right corner is mostly straight, the arrival rate of new cases is constant.



2. Change in the global process

Approach I used: As mentioned before I changed sorting to "Sort on time: timestamp of first events and Color Attribute to "C: Activity classifier". What I saw: I saw that the colored dots representing different activities changed around June 2011 from dots in various colors indicating different starting activities to a more homogeneous pattern starting with orange dots.

My analysis: As the pattern of the colored dots changed around June 2011, I conclude that the activity sequence in the process was changed such that cases were started with activity T10 Determine necessity to stop indication - completely for most subsequent cases.

Approach I used:

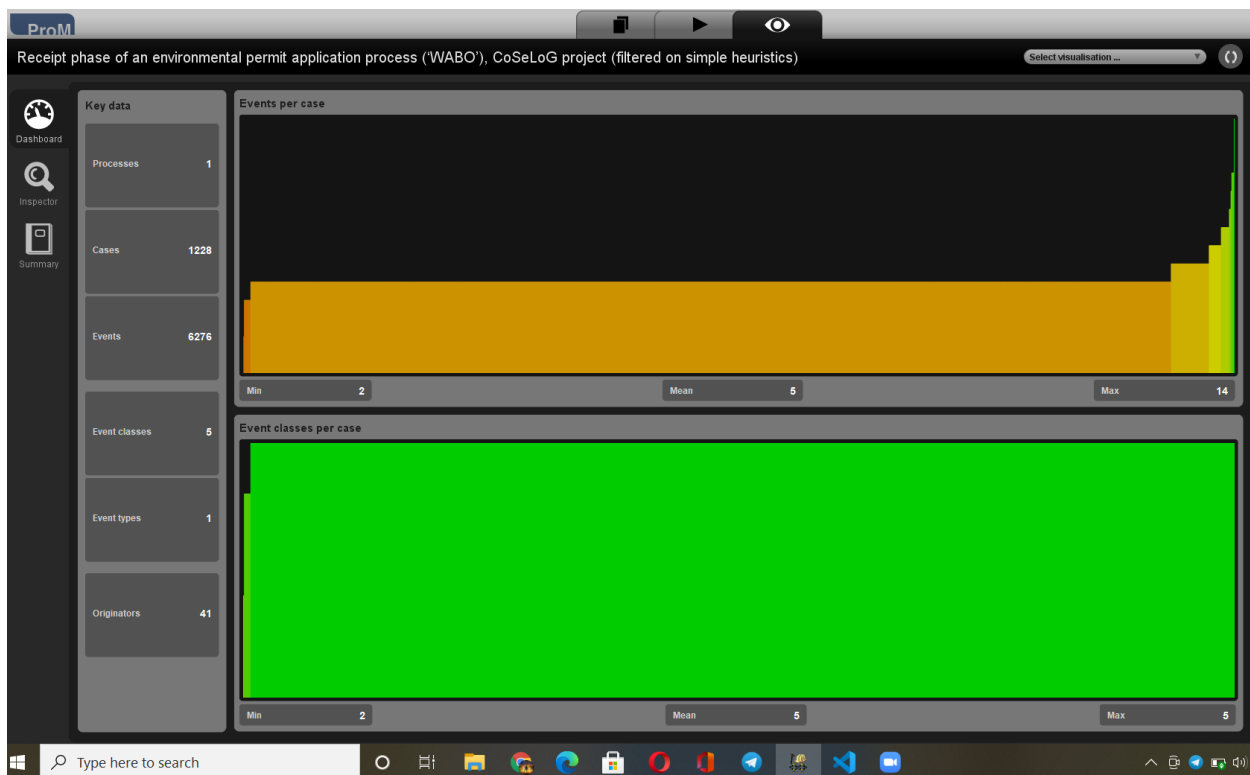
1. Settings I used below settings: - Percentage: 80% Start activity: Confirmation of receipt complete - End activities: T05 and T10 - Activities: Confirmation of receipt complete, T02, T04, T05, T06, T10 (main activity sequence)

2. Motivation I mostly used the standard settings as suggested when using the simple heuristics filter, i.e. - Select top 80% in line with the pareto principle Start case End Events: Use T05 Print and send

Events: Use 'Confirmation of receipt complete' as Start event as it is the starting point of the process in any confirmation of receipt - complete' or T10 Determine necessity to stop indication - complete' as End events as they were the most frequent activities at the end of the process - Event Filter: Use all activities from the main path 'Confirmation of receipt complete -> 'T02 Print and send confirmation of receipt - complete' -> T06 Determine necessity of stop advice - complete T10 Determine necessity to stop indication - complete'.

Check confirmation of receipt - complete >> T04 Determine confirmation of receipt - complete' - T05

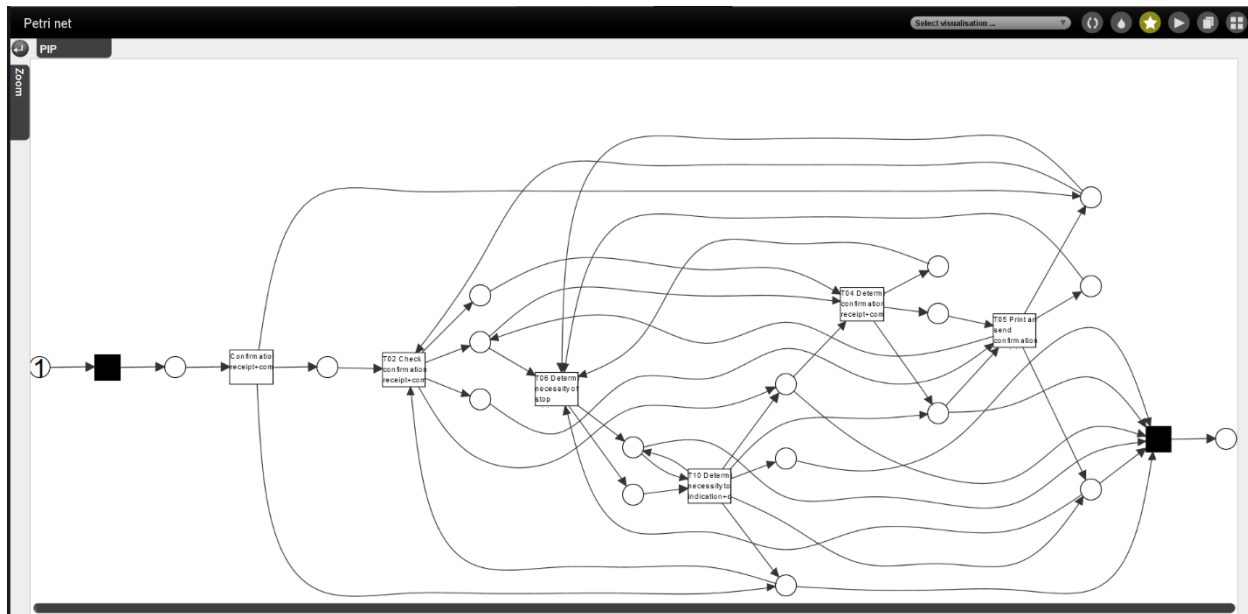
3. Plug ins used I first used the Alpha Miner with standard settings on the filtered event log with the standard settings I then used the Inductive Miner with standard



settings.

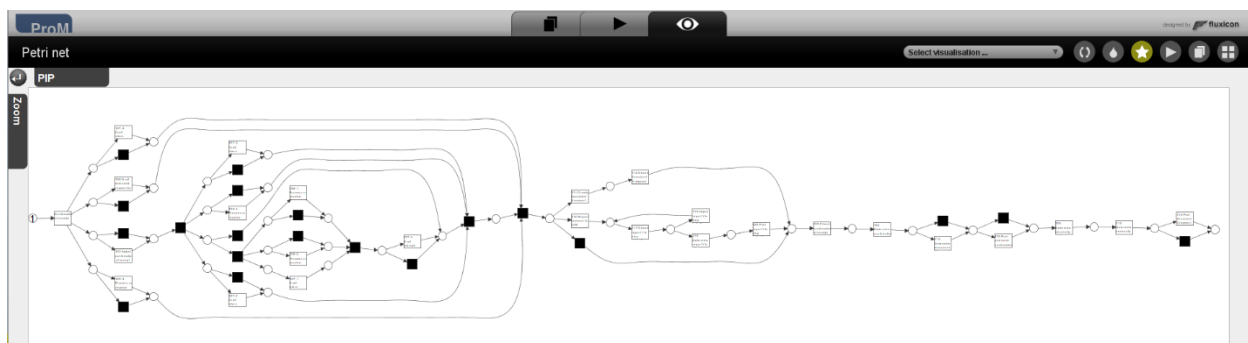
What I saw: The Petri net produced by Alpha Miner had only T04, T05 and T10 connected. The starting activity 'Confirmation of receipt, T02' and T06 were shown as single unconnected activities. See screenshot

The Petri net produced with Inductive Miner was clearly readable and showed a consistent Petri net. See screenshot



4. Main process and notable parts the main process as shown in the Petri net derived by using Inductive Miner shows that there are two concurrent streams in the process after the starting point 'Confirmation of receipt. In the first stream activities T06 Determine necessity of stop advice - complete' and T10

Determine necessity to stop indication - complete have to be completed in this order. In the second stream activities to check on T02 Check confirmation of receipt - complete! T04 Determine confirmation of receipt - complete and T05 Print and send confirmation of receipt - complete' have to be conducted one after the other. The process can be finished after T10 and T05 are finished.



My analysis: I concluded that Alpha Miner is not able to detect the correct Petri net due to its known weaknesses. I could see that Inductive Miner is much more robust with respect to noisy and incomplete data and showed a consistent Petri net.

Normative model: conformance checking on process model and on full unfiltered original event log

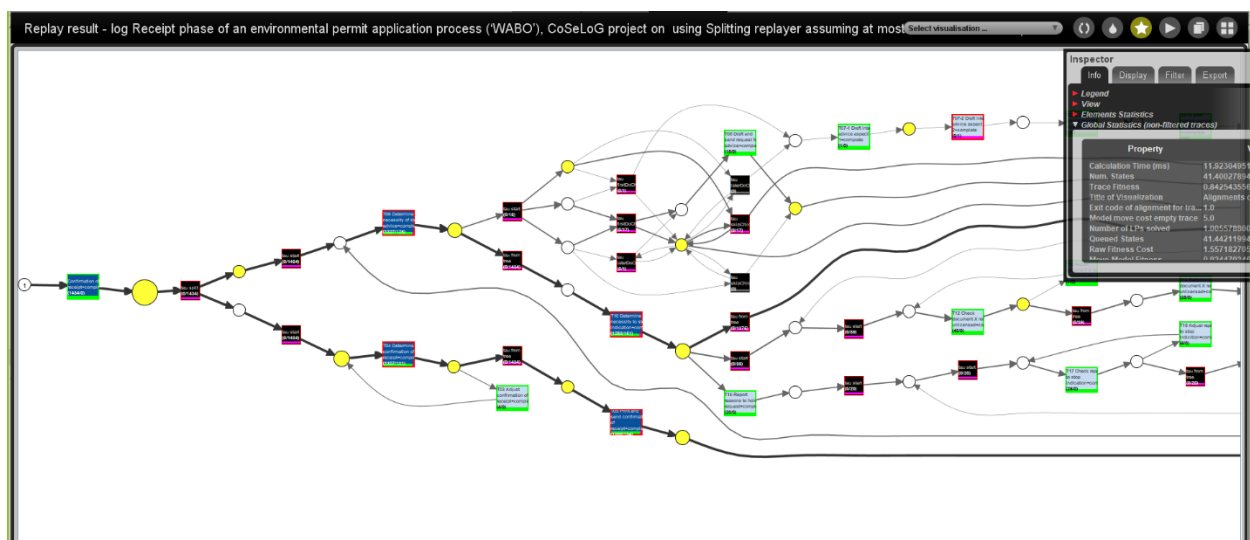
Approach used:

I followed the instructions to align the process model with the event log in order to show the normative model with the respective conformance information. I zoomed in on the part with the most deviations shown by the big yellow circles.

What I saw:

1. I saw that highest deviations occur with respect to activities T04' or 'T06' but also to a lower degree for 'T05' and T10', This is shown in the visualization

The replay fitness or trace fitness can be seen when selecting 'Global Statistics' in the 'Inspector' and is 0.8425 rounded to 4 decimal places.



The element statistics of T06 Determine necessity of stop advice complete can be found by clicking on activity 'T06 and 'Element Statistic in the 'Inspector'. 1'327 log and model moves were consistent in 125 cases there were moves in model only.

4. This results in around 10% being incorrect which is rather significant

My analysis:

The high deviations with respect to 'T04', 'T05', 'T06 and 'T10' should be investigated further in order to find out whether the model or the event log are wrong. Furthermore it has to be evaluated whether the deviations are good or bad in order to impose the right measures. For T06 the error rate of 10% is significant and should be investigated in any case.

Approach I used:

After using plug-in 'Mine for a Subcontracting Social Network' according to the instructions given above I set the slider "Edges removed" somewhat more to the right in order to filter out weaker connections. In addition to that I selected 'Group Clusters in order to identify groups.

I then used the Dotted Chart visualizer and changed the Y Axis Attribute' to 'C: Resource classifier' and the color attribute to 'C: Activity Classifier'.

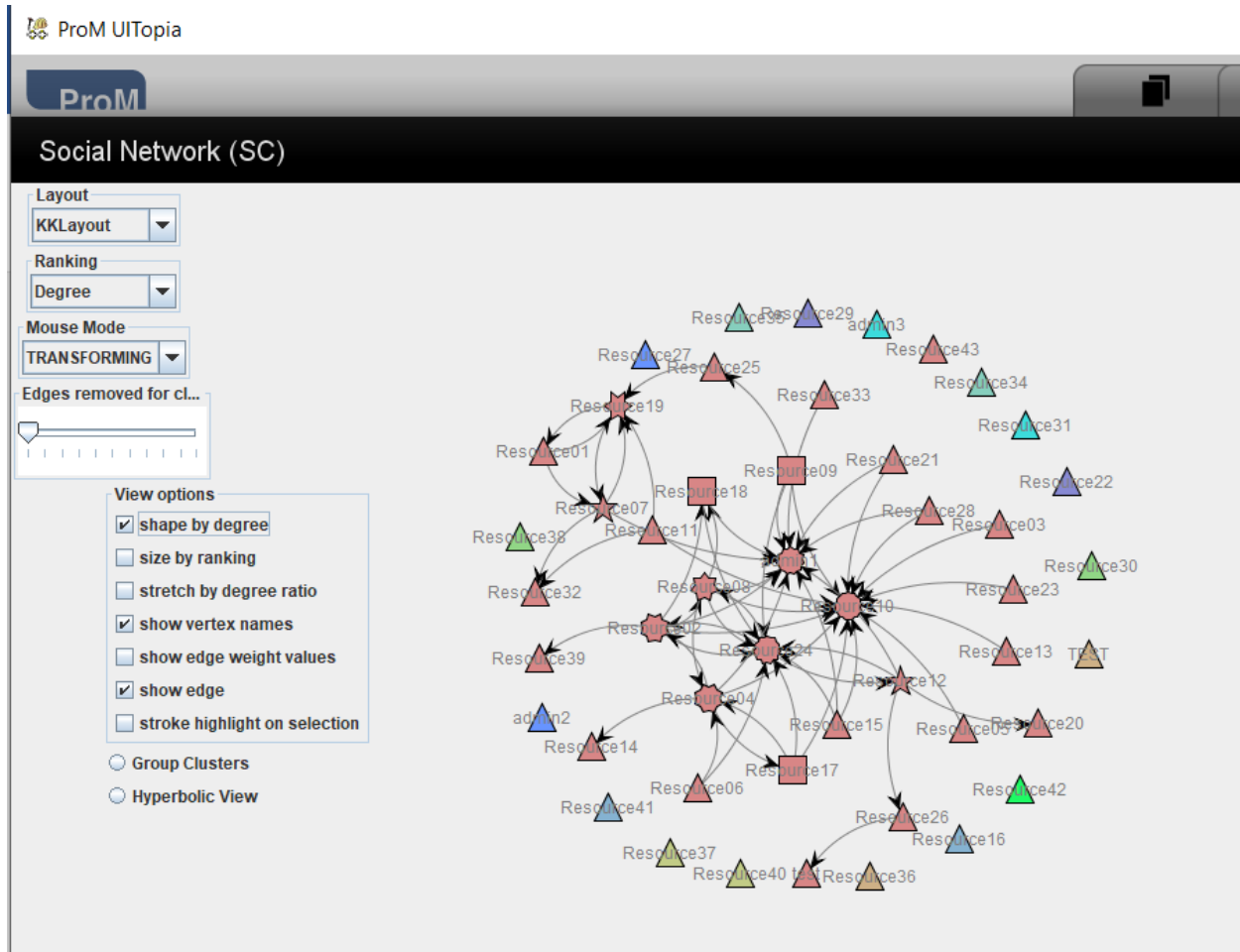
What I saw:

1. I identified two groups and apart from that individual resources. Group 1 consists of admin 1, Resource 02, 03, 04, 08, 09, 10, 12, 15, 17, 18 and 24. The comparatively smaller group 2 consists of Resource 01, 07, 11 and 19. Most resources belong to the 'Subcontracting network whereas the others are disconnected.

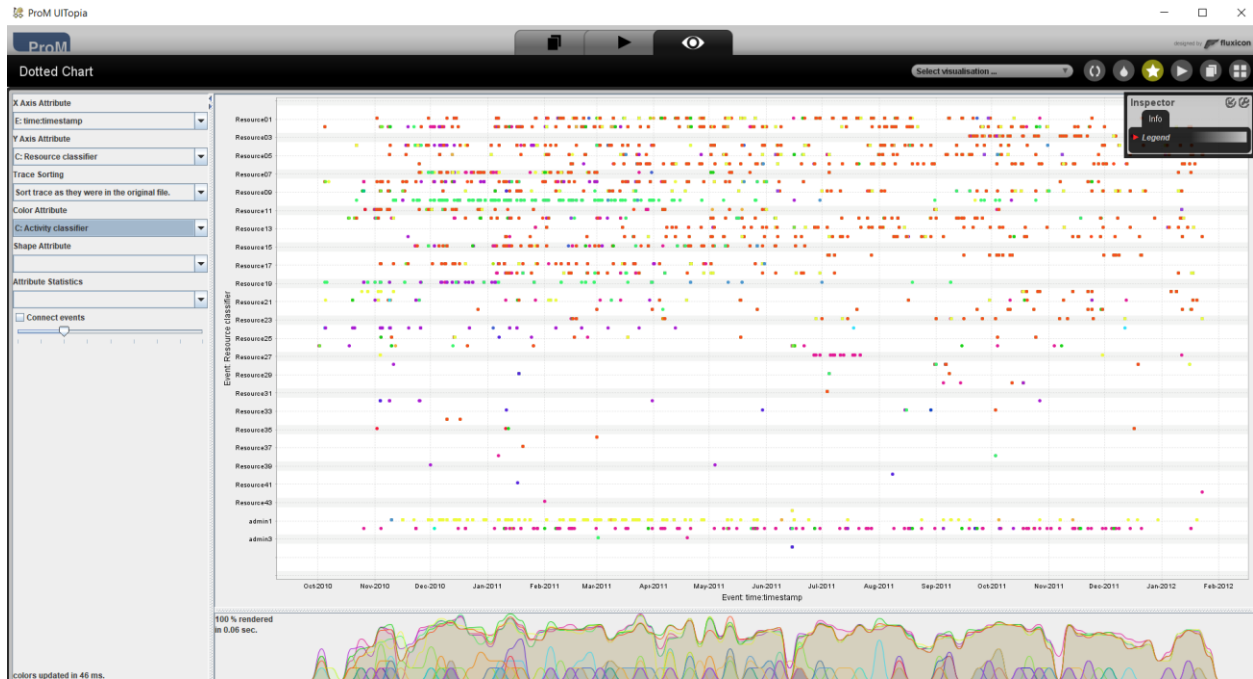
3. Not all users execute activities from the start of the event log. Various resources join later, e.g., Resource 38, 41 or 43. See screenshot

Some users like 'admin 1' execute only particular activities. Many other resources execute various different activities.

My analysis:



It looks like some of the activities can only be conducted by a few specific resources which might keep them very busy and result in potential bottlenecks in the process.



Three main observations:

1. Many cases in the event log contain loops, in most of the cases loops back to previous activities but also loops in itself indicating more instances of an activity. Some of them show a very low performance and should be analyzed accordingly.
2. Some of the activities have a relatively high error rate such as T06, which is not correctly completed in 10% of the cases. This requires attention in order to derive the right conclusions for process optimization.
3. Some activities are carried out only by a few resources as the analysis of the Dotted Chart with respect to resources and activities shows. Training more workers on those tasks might speed up the process.