Question-1:

Open the event log ('Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project.fbt') in Disco and switch to the 'Statistics' view.
Without switching to other views, use the statistics view to answer the following three subquestions:

1. How many events are there on average per case?
2. Can you indicate whether each case seems to be unique or whether many cases follow the same activity sequence?
3. What is the main observation that can be made from the 'Events over time' graph?

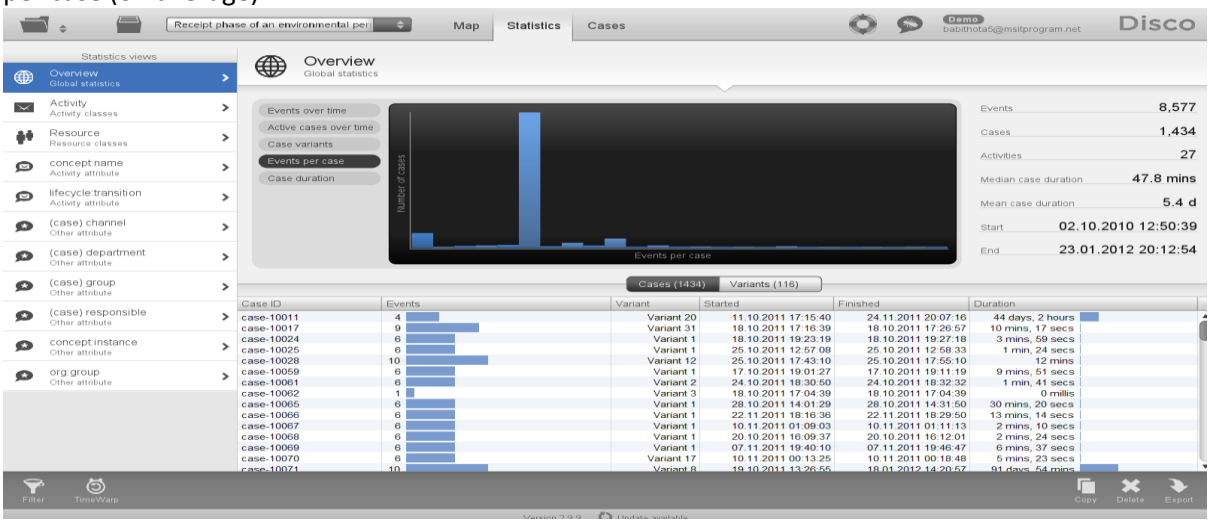
Solution:

Process:

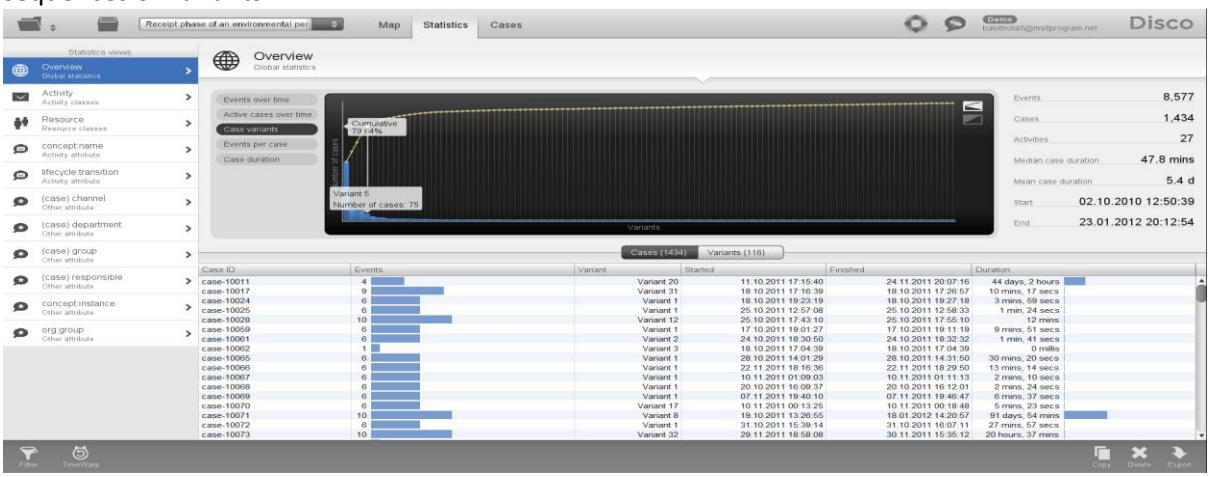
Open Disco
Click “Open file” button in the “Load your own data” option of the welcome screen.
Look for the 'Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project.fbt' and should import it.
From the three buttons centred on the top of the screen, click the “Statistics” one to switch to the “Statistics” view.

Observance

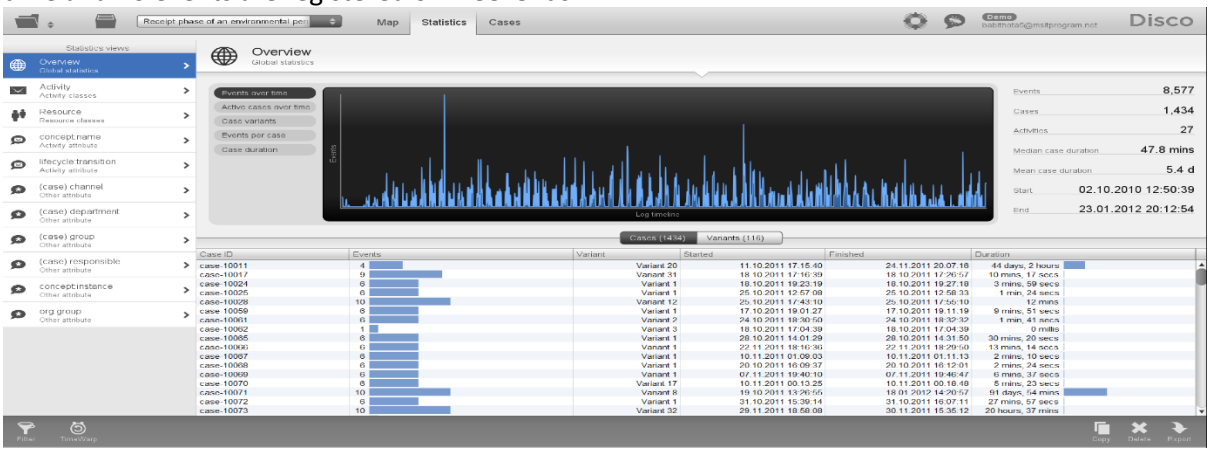
1. Click the “Events per case” option. We can see there are approximately 6 events per case. If we divide total number of events (8577) by the number of cases (1434), we have 5.98 events per case (on average).



2. Click the "Case variants" option. Here, each case is not unique, there are many cases following the same activity sequence. Here, 1434 cases follow 116 different activity sequences or variants.



3. Click the "Events over time" option. We might conclude that events are very distributed over time and no events are registered on weekends.



Question-2

While still in Disco, switch to the 'map' view to display a process map.
Using the map view, change the activity and path detail settings in order to create a comprehensible process map (e.g., a process map that could be printed on one A4 or letter paper or shown on a single computer screen while still being readable in full).

1. Discuss this process map, what is the main process?
2. Which activities and paths between activities are frequent?

In your answer, include the settings you used for both the activity and path sliders.

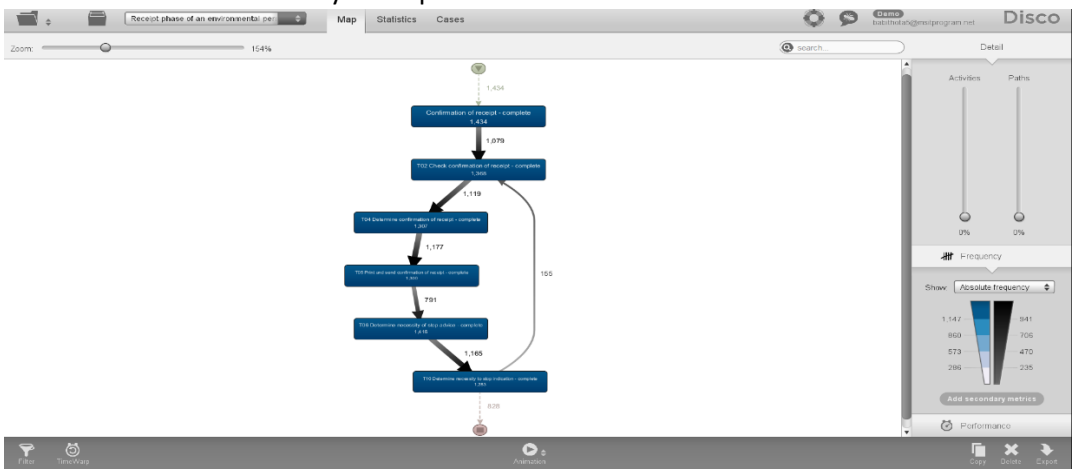
Solution:

Process

Click "Maps" button to load the Maps view
Using the "Activities" and "Paths" sliders in the right part of the screen, I had set a 0% threshold for both elements to create a comprehensible process map

Observance

The main sequence of activities follows:
1. Confirmation of receipt
2. T02 – Check confirmation of receipt
3. T04 - Determine confirmation of receipt
4. T05 - Print and send confirmation of receipt
5. T06 - Determine necessity of stop advice
6. T10 - Determine necessity to stop indication



Question-3

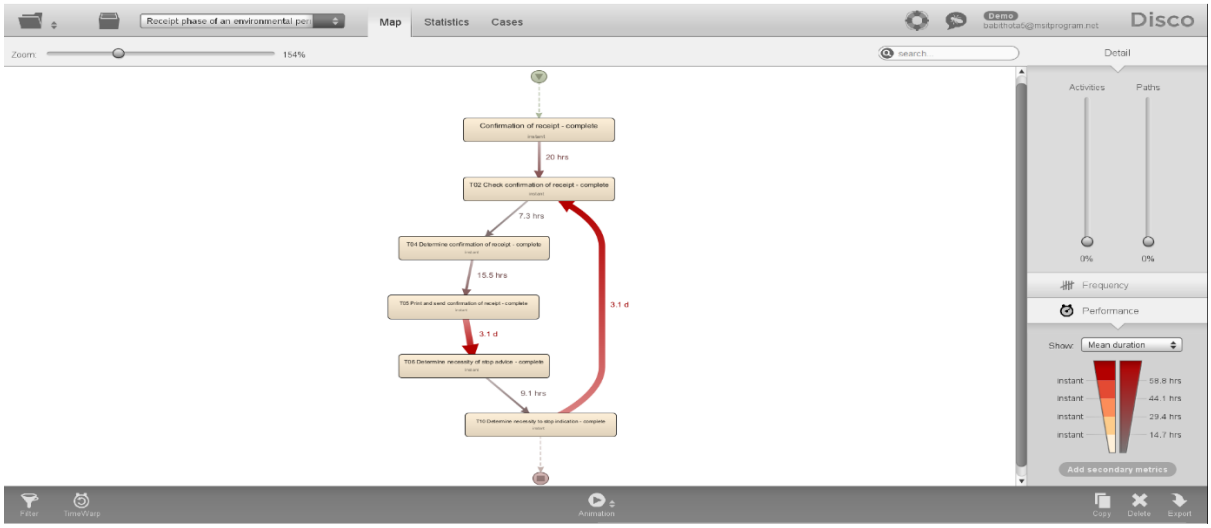
While still in Disco, and while using the same process map (e.g., do not change the activity and path settings), switch to the performance projection.
Discuss where the process takes most time, e.g., where there are possibilities for improvement. Relate these times (of the bottlenecks) to the time spent in other parts of the process. In other words, discuss how severe the bottleneck is with respect to the time spent on other activities.
Also explicitly mention the performance metric chosen (e.g., total, mean, median, or max) and why you have chosen this setting.

Solution:

Process

Click over the "Performance" control to show the performance projection.
Next, select the "Mean duration" in the "Show" drop-down menu to have an overview of which the main problems in the process are.

Observance



I have chosen mean duration for the effective results as the mean values will be accurate.

Question-4

Now load the original event log in Prom. Visualize the event log using the Dotted Chart or XDottedChart visualizer (by pressing the 'eye'-icon with the event log selected and switching to the Dotted Chart or XDottedChart visualizer).
Using the Dotted Chart, answer the following questions:

1. Is the arrival rate of new cases constant? If not, when are there fluctuations? If yes, how can we see this from the Dotted Chart?
2. Can you observe a change in the global process?

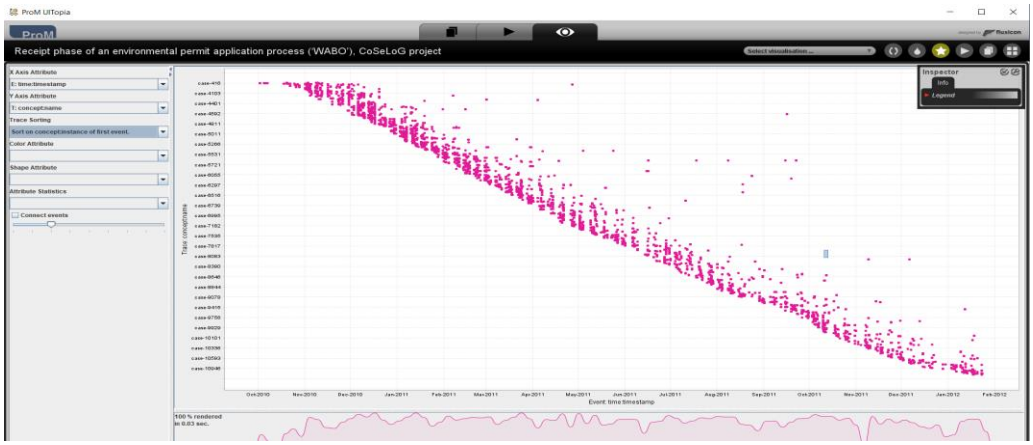
Note that you don't need to change the component, time or colouring settings. You can however re-sort the traces on the time of the first event, and zoom in or out if you want.
The Dotted Chart is explained in lecture 4.8: 'Exploring Event Data'.

Solution:

Process:

Extract the "Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project.gz" to get a "Receipt phase of an environmental permit application process (_WABO_) CoSeLoG project.xes" file.

Open Prom and import the extracted file.
Click "Accept" in the "Select import plugin" window that opens to use the ProM log files Import Plugin
Click the "Eye" button to open the dashboard that allows you to visually the data collected in the event log
In the "Select visualization" drop-down menu, select the "Dotted Chart" one
Here is the resultant Dotted Chart



Observance:
As illustrated by the figure above, the arriving of new cases is constant since we have sorted the traces according to the instance of the first event, and we cannot find abrupt jumps between them. We might notice a lower flow of cases at the beginning, but it is negligible. Regarding the global process, by looking at the activities performed along the timeline, we can conclude that different activities were performed during the first part of the timeframe corresponding to the event log.

Question-5

You are now asked to discover a Petri net on the event log. However, the unfiltered event log results in an incomprehensible Petri net. Therefore, you are allowed to run the 'Filter log using simple heuristics' plug-in *once* on the original event log to discover a Petri net on the filtered event log.

1. Clearly indicate which settings you have used for the 'Filter log using simple heuristics' plug-in.
2. Explicitly motivate the filtering settings chosen, why did you pick this percentage or selection of activities?
3. Discuss and argue which plug-in (or chain of plug-ins) you have used to discover a Petri net, for instance by comparing two or more plug-in results and arguing why one of the Petri nets is better.
4. Explain the (best) Petri net: what is the main process and what are notable parts of the Petri net?

Note that this question requires you to experiment with different filtering settings and discovery plug-ins. You are not required to describe *everything* you have tried but found unsuccessful. Only describe the successful combination of plug-ins and its result(s) and argue why your final result is 'good'.

Suggested list of plug-ins or plug-in chains to produce a Petri net:

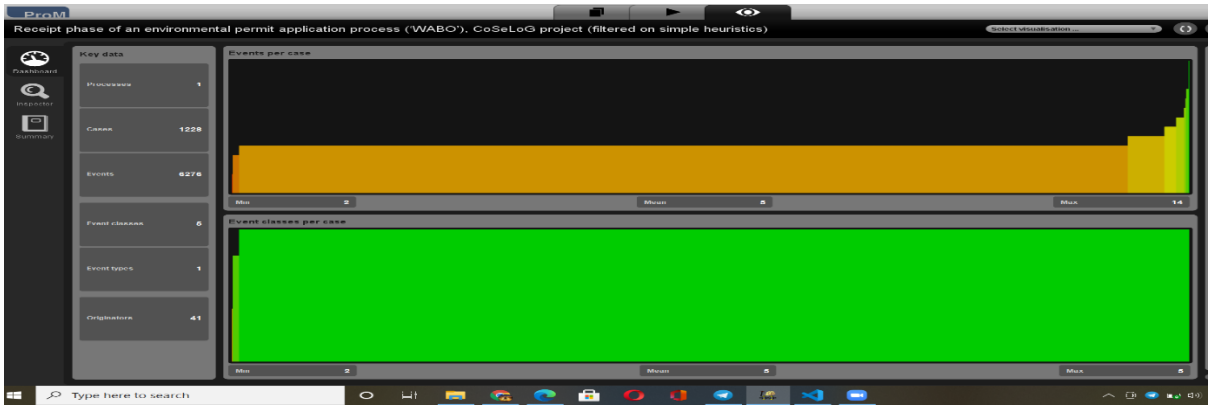
- Mine for a Petri Net using Alpha-algorithm
- Mine for a Petri Net using ILP
- Mine for a Heuristics Net using Heuristics Miner *followed by* Convert Heuristics net into Petri net

- Mine for a Petri net with Inductive Miner

Solution:

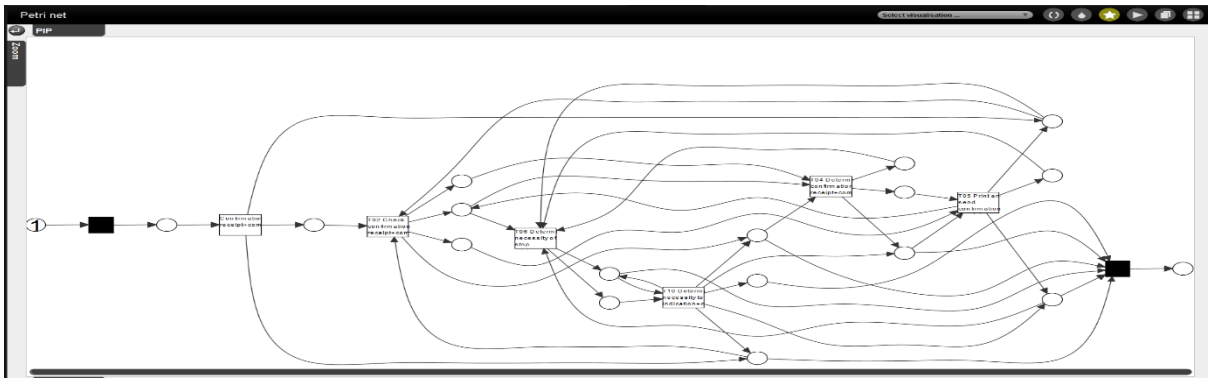
Process

Filtered the Log using the "'Filter log using simple heuristics' plug-in". The different thresholds were set at 80% since it happens to be still acceptable number of activities, so the main process is supposed to be still present in the result.

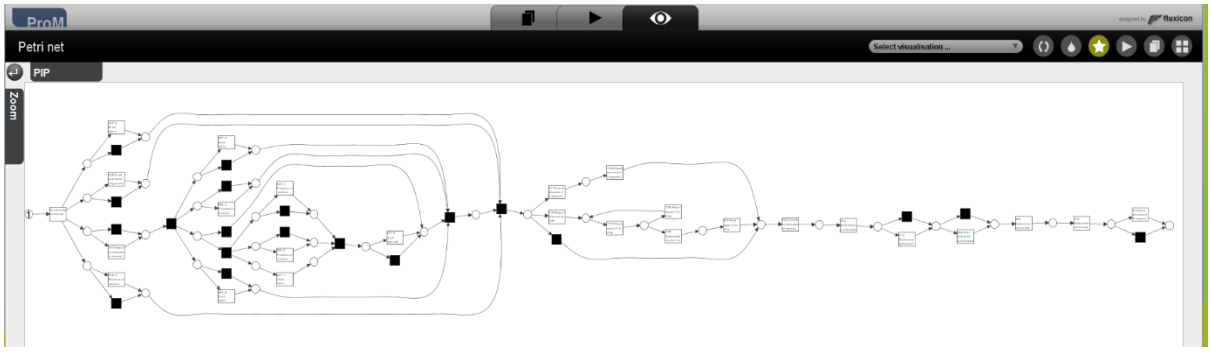


Mined the filtered log and generated Petri Net using the " Mine for a Petri Net using ILP" and “Mine for a Petri net with Inductive Miner” plugins.

Petri Net using ILP



Petri Net using Inductive Miner



Observance

The configuration finally used has to do with the results obtained using the rest of configurations suggested: the first two options resulted in Petri Nets with a high number of isolated places. Besides, not filtering the log before mining it to produce the net resulted in a Petri Net with a huge number of places which hampered drastically any kind if analysis. By contrast, the resulting Petri Net serves to illustrate clearly the main process, which happens to be very similar to the one already discovered in Disco (and discussed above) when using appropriate levels of filtering by using the corresponding sliders.

Question-6

The organization has a process model that describes the 'should be' process (i.e., a normative process model). Load the file 'normativeModel.pnml' into Prom and apply conformance checking on this process model, and on the full unfiltered original event log.

1. Include a screenshot of the part of the normative process model, with the conformance information projected onto it, that shows where most of the deviations occur.
2. What is the replay fitness (the 'trace fitness' statistic) of the event log on the normative process model?
3. Select the transition 'T06 Determine necessity of stop advice+complete' (on the top left of the model) and discuss its element statistics: how many times is the transition executed correctly and how many times incorrectly?
4. Using the element statistics of transition 'T06 Determine necessity of stop advice+complete', what can you say about the (in)correct execution of this activity?

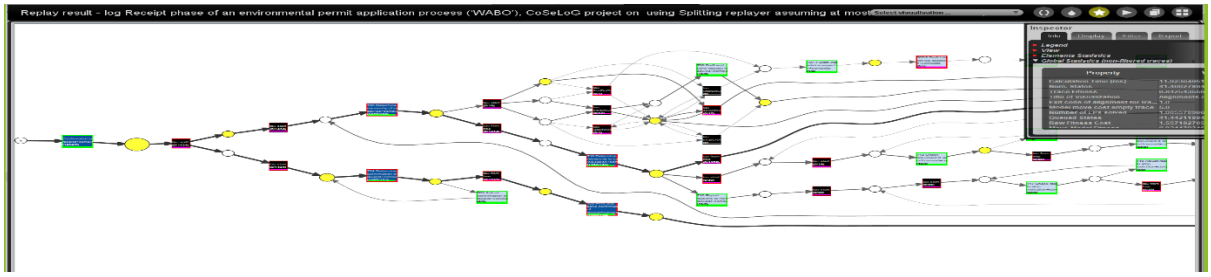
Instructions to align the process model with the event log:

1. Import the normative model using the 'PNML Petri net files' importer.
2. Select the imported normative Petri net and the event log, start the plug-in called ‘Replay a Log on Petri Net for Conformance Analysis’ (not the variant with performance!), and click 'yes' in the 'No Final Marking' pop-up.
3. Select the 'sink' place on the left (note: do not select '0-sink' etc.) and click the button 'Add Place >>' to add the place 'sink' to the candidate final marking list. Now click 'Finish'.
4. Click 'Finish' in the mapping wizard.
5. Click 'No, I've mapped all necessary event classes' to indicate that some events are not present in the normative model.
6. Now click 'Next' and 'Finish'. The normative process model is shown with conformance information projected onto it.

If you followed these instructions exactly you do not need to mention these steps in your answer. More information regarding this conformance technique is provided in lecture 4.7: 'Aligning observed and modelled behaviour' (and to a lesser extend in the lectures 4.3 through 4.6).

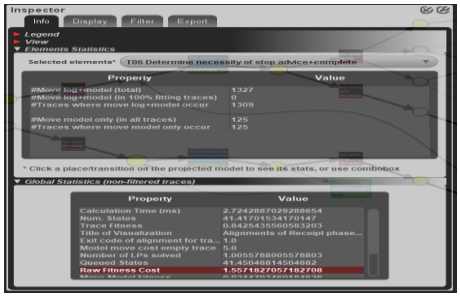
Solution:

Process



Observance

The above screenshot shows the left-part of the normative process model, where most deviations occur, as indicated by the yellow places.



As the Inspector view shows, the replay fitness (the 'trace fitness' statistic) of the event log on the normative process model is 0.84
Next, the element statistics for the transition T06
From the above screenshot, the transition T06 is correctly executed 1327 times whereas it is wrongly executed 125 times.
When compared with the remaining transitions presenting problems, the number of wrong executions is more or less. All these transitions be revised to identify the errors made at the moment of deriving a model since the model is not in line with the observations in the log in around 10% of the cases.

Question-7

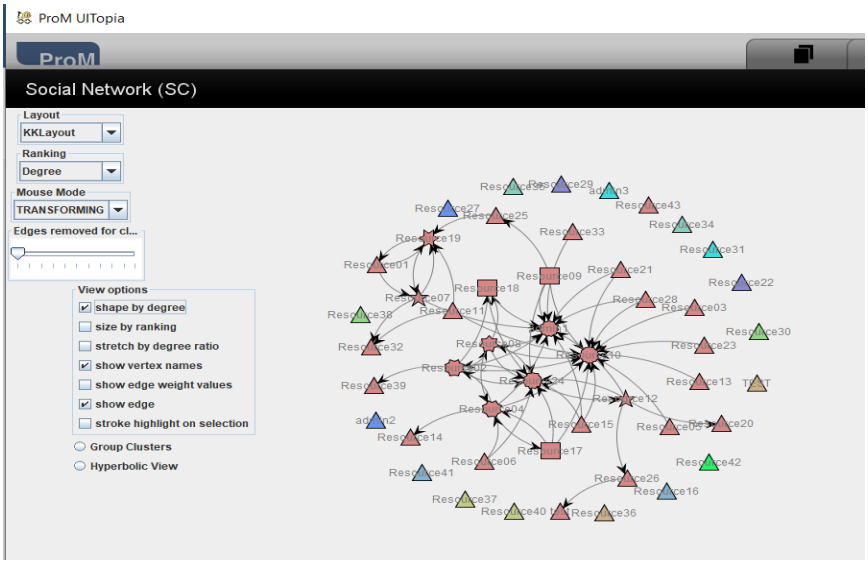
The final analysis you have to perform on the original event log is a resource analysis, e.g., looking at the user behaviour in the event log.

1. Use the plug-in 'Mine for a Subcontracting Social Network'. Note that subcontracting means that if individual *j* frequently executed an activity in-between two activities executed by individual *i*, then individual *i* subcontracted work to individual *j*. Answer the following question using this view: Can two or more groups of users be distinguished? Explicitly discuss the settings you have used in the resulting visualization.
2. Again, use one of the two Dotted Chart plug-ins. For the XDottedChart change the component type to 'org:resource'. If you use the Dotted Chart visualizer change the 'Y Axis Attribute' to 'C: Resource classifier' and the colour attribute to 'C: Activity Classifier'. Answer the following two questions using this view:
3. Are all users executing activities from the start of the event log, or are some users joining later?
4. Are users mainly executing particular activities or are most users executing most of the activities?

Solution:

Process

Set the original event log as input and Click Play button and select the "Mine for a Subcontracting Social Network' plugin. Then click 'Start'
As a result, the following graph is obtained



Back to the 'Resources' window, select the original event log and click the 'Eye' button to load the visualization of the traces. Next, Choose the Dotted Chart (Log Projection) from the drop-down

menu ‘Select visualization’. In the dotted graph obtained, we change the 'Y Axis Attribute' to 'C: Resource classifier' and the colour attribute to 'C: Activity Classifier'. As a result, we obtain the following dotted graph.

Observance

From the Social Network graph shown above, by the resulting concentric shape, we can distinguish two groups of resources: one group is those located in the centre of the graph, with a relevant number of connections and the other group is with the group of resources surrounding them, which happens to be isolated. To better illustrate this observation, note that we have activated the "shape by degree" option which helps on the identification of the two different groups of users, gearwheels vs squares and triangles.

Are all users executing activities from the start of the event log, or are some users joining later?

Even though many of the resources start working early in the timeframe corresponding to the event log, there are some resources that start executing activities very late

Are users mainly executing particular activities or are most users executing most of the activities?

We can conclude that some users are executing a wide variety of activities, while others are focused on some specific types of activities.

Question-8

To conclude this assignment, briefly discuss three observations you have made during your analysis that you would like to communicate to the business user.

Think for instance of possible improvement opportunities and starting points for further investigation.

Solution:

1. There are some activities that have been detected as possible bottlenecks and should be therefore inspected in detail.
2. Some deviations between the normative process model and the actual execution of the process has been detected. These deviations should be inspected as well.
3. There are some resources apparently overloaded, or at least receiving much more work that other resources.