



UNIVERSITÀ DELLA CALABRIA

DIPARTIMENTO DI  
INGEGNERIA INFORMATICA,  
MODELLISTICA, ELETTRONICA  
E SISTEMISTICA

DIMES

Corso di Laurea Magistrale in  
Ingegneria Informatica

Progetto di “Metodi Informatici per l’analisi dei processi”

Relazione BPI-Challenge 2012

**Docenti**

Prof.ssa Antonella Guzzo  
Ing. Eugenio Vocaturo

**Studenti**

Antonio Marino matr. 224598  
Aron Raffaele matr. 224437  
Mattia Costabile matr. 224827

Anno Accademico 2020/2021

# Indice

1	Introduzione .....	5
2	Descrizione del Dataset .....	5
2.1	Mappa generale del processo .....	8
3	Analisi statistica degli eventi .....	11
3.1	Major behaviour e varianti .....	13
3.2	Dotted Chart .....	15
4	Process Discovery.....	17
5	Process Conformance .....	20
5.1	Conformance sottoprocessi A, O e W .....	21
5.2	Conformance Celonis .....	22
5.3	Considerazioni sulle varianti .....	23
6	Performance Analysis.....	24
6.1	Performance sottoprocessi A, O e W .....	25
6.2	Performance Celonis .....	28
7	Resource analysis.....	31
7.1	Mine for Handover-of-Work Social Network .....	31
7.2	Mine for a Reassignment Social Network.....	34
7.3	Mine for a Similar-Task Social Network .....	35
7.4	Mine for a Subcontracting Social Network.....	36
7.5	Mine for a Working-Together Social Network .....	37
7.6	Dotted chart .....	38
7.7	Analisi risorsa 112 (Celonis).....	41
7.8	Analisi risorsa senza nome (Celonis) .....	42
8	Conclusioni .....	43

# Indice Figure

Figura 1: Panoramica generale (ProM) .....	5
Figura 2: Eventi iniziali e finali (ProM).....	7
Figura 3: Attività più frequenti (Celonis) .....	7
Figura 4: Mappa generale del processo (Disco).....	8
Figura 5: Mappa esplosa del sotto-processo "A" (Disco) .....	9
Figura 6: Mappa esplosa del sotto-processo "O" (Disco) .....	9
Figura 7: Mappa esplosa del sotto-processo "W" (Disco) .....	10
Figura 8: Top 10 richieste di prestito/scoperto (Disco) .....	10
Figura 9: Top 10 richieste di prestito/scoperto anomale (Disco) .....	10
Figura 10: Panoramica statistiche eventi nel tempo (Disco).....	11
Figura 11: Panoramica statistiche dei case nel tempo (Disco).....	11
Figura 12: Panoramica frequenze attività (Disco) .....	12
Figura 13: Panoramica frequenza risorse (Disco).....	12
Figura 14: Happy path (Celonis).....	13
Figura 15: Top 10 varianti (Disco) .....	13
Figura 16: Dotted chart case/timestamp (ProM).....	15
Figura 17: Dotted Chart eventi/timestamp (ProM) .....	16
Figura 18: Modello estratto da Alpha miner (ProM) .....	17
Figura 19: Modello estratto da Alpha miner + "filter log using simple heuristics" (ProM) .....	18
Figura 20: Modello estratto da Heuristic miner + "convert heuristics net into petri net" (ProM) .....	19
Figura 21: Modello estratto da Inductive miner con Noise threshold al 20% (ProM).....	19
Figura 22: "Replay a Log on Petri Net for Conformance Analysis applicaton" su Petri net ricavata da Inductive miner (ProM) .....	20
Figura 23: Conformance sublog A (ProM) .....	21
Figura 24: Conformance sublog O (ProM) .....	21
Figura 25: Conformance sublog W (ProM) .....	22
Figura 26: Modello "to be" (Celonis).....	22
Figura 27: Conformance (Celonis) .....	23
Figura 28: Conformance con modifiche nell'allowlist (Celonis) .....	23
Figura 29: Performance mappa generale (Disco) .....	24
Figura 30: Performance sublog A (Disco) .....	25
Figura 31: Performance sublog O (Disco) .....	26
Figura 32: Performance sublog W (Disco) .....	27
Figura 33: Throughput generale (Celonis).....	28
Figura 34: Bottleneck (Celonis).....	29
Figura 35: Throughput con "W_Completeren_aanvraag -> W_Nabellen_offertes" (Celonis) .....	29
Figura 36: Throughput con "W_Nabellen offertes ->W_Validatoren aanvraag" .....	30
Figura 37: Throughput con "W_Nabellen offertes -> W_Nabellen offertes" .....	30
Figura 38: Panoramica frequenza delle risorse (Disco) .....	31
Figura 39: Handover of Work (ProM) .....	31
Figura 40: Handover of Work + skretch by degree ratio (ProM) .....	32

Figura 41: Handover of Work + sketch by degree ratio + size by ranking (ProM).....	33
Figura 42: Mine for a Reassignment Social Network (ProM) .....	34
Figura 43: Mine for a Similar-Task Social Network (ProM).....	35
Figura 44: Mine for a Subcontracting Social Network (ProM) .....	36
Figura 45: Mine for Working-Together Social Network .....	37
Figura 46: Dotted Chart delle risorse nel tempo (ProM) .....	39
Figura 47: Dotted chart delle risorse rispetto alle attività (ProM) .....	40
Figura 48: Numero degli eventi della risorsa 112 (Celonis) .....	41
Figura 49: Distribuzione giornaliera degli eventi della risorsa 112 (Celonis) .....	41
Figura 50: Numero degli eventi della risorsa NULL (Celonis) .....	42
Figura 51: Distribuzione giornaliera degli eventi della risorsa NULL (Celonis) .....	42

# 1 Introduzione

Lo scopo di questo progetto è stato quello di analizzare il dataset della *BPI Challenge 2012* relativo alle richieste di prestito e scoperto presso un istituto finanziario olandese. In particolare, è stata effettuata dapprima un'analisi statistica degli eventi e in seguito sono stati sviluppati i punti riguardanti il *process discovery* e relativa *conformance* dei modelli estratti, l'analisi delle performance e delle risorse. Infine, è stata data risposta ad alcune domande inerenti a specifiche del progetto. L'event log di partenza, in seguito ad alcune considerazioni, è stato suddiviso in tre *sublog* distinti, ovvero “A” “O” e “W”, proseguendo l'analisi separatamente ci si è accorti che non sempre ha messo in luce criticità che invece sono emerse con un'analisi più generale. Per effettuare le analisi sono stati utilizzati in parallelo i tool *ProM*, *Disco* e *Celonis*. Per concludere, sono state sintetizzate le scoperte più interessanti e le principali criticità del processo.

# 2 Descrizione del Dataset

Il processo analizzato nell'event log preso in analisi descrive la richiesta di prestiti personali o scoperti, ovvero prelievi di somme che eccedono in conto corrente. I dati collezionati coprono circa 165 giorni (dal 01.10.2011 al 14.03.2012), registrando una totalità di 262.000 eventi in 13.087 case, ognuno dei quali rappresenta una richiesta di prestito o scoperto (la cui cifra è indicata con l'attributo *AMOUNT\_REQ*) da parte del cliente. Il numero di risorse presenti all'interno del processo è pari a 69.

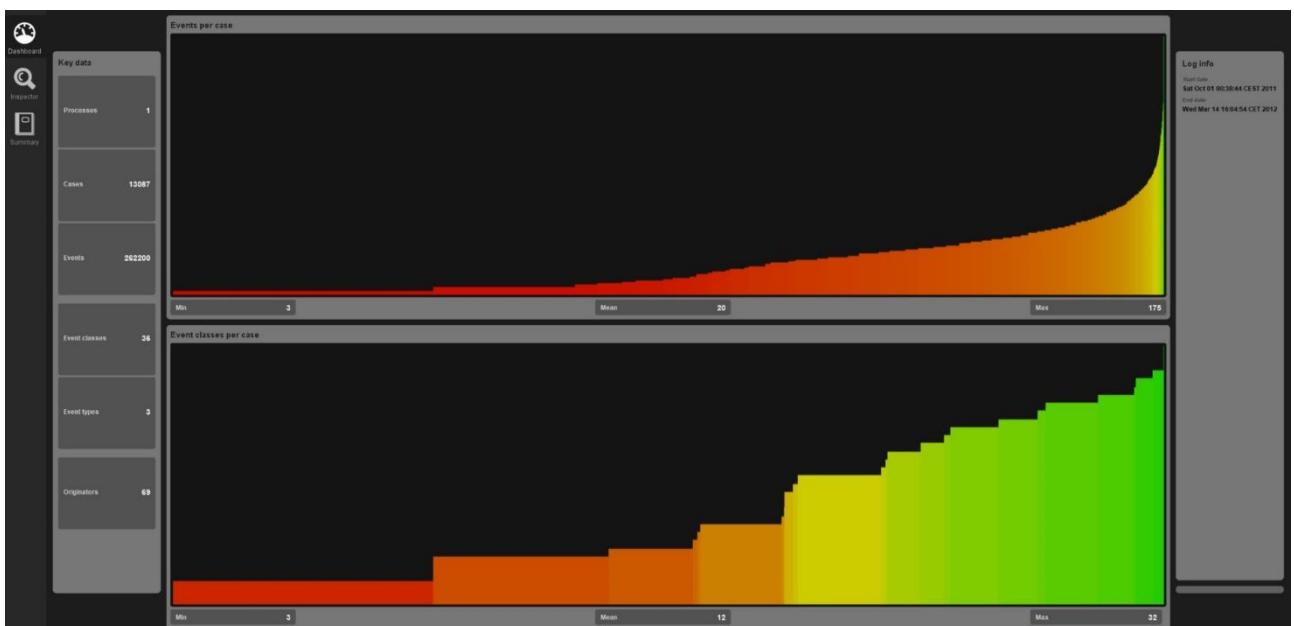


Figura 1: Panoramica generale (*ProM*)

Da come si evince dal sito web della *BPI Challenge* presa in carico, il registro eventi è una fusione di tre processi secondari intrecciati che vengono distinti in base ad un carattere iniziale “A\_”, “O\_” e “W\_”; inoltre, per ciascuna tipologia, vi è un’ulteriore suddivisione in base al ciclo di vita del processo.

Quanto detto viene meglio specificato nella tabella sottostante:

<b>Tipologia attività (carattere iniziale)</b>	<b>Descrizione</b>
“A_”	<p>Rappresentano gli stati delle richieste fatte dai clienti.</p> <ul style="list-style-type: none"> <li>• <i>A_SUBMITTED / A_PARTLYSUBMITTED</i> = presentazione della richiesta inviata dal cliente;</li> <li>• <i>A_PREACCEPTED</i> = accettazione della richiesta con riserva, in attesa di ulteriori informazioni;</li> <li>• <i>A_ACCEPTED</i> = domanda accettata in attesa di essere revisionata per il completamento;</li> <li>• <i>A_FINALIZED</i> = richiesta ultimata;</li> <li>• <i>A_APPROVED / A_REGISTERED / A_ACTIVATED</i> = stati finali di successo per la richiesta;</li> <li>• <i>A_CANCELLED/ A_DECLINED</i> = stati finali di insuccesso per la richiesta.</li> </ul>
“O_”	<p>Rappresentano gli stati delle comunicazioni di offerte fatte ai clienti.</p> <ul style="list-style-type: none"> <li>• <i>O_SELECTED</i> = selezione del cliente che riceve l’offerta;</li> <li>• <i>O_CREATED</i> = creazione dell’offerta da inviare al cliente;</li> <li>• <i>O_PREPARED / O_SENT</i> = preparazione dell’offerta e invio della stessa al cliente;</li> <li>• <i>O_SENT BACK</i> = ricezione della risposta del cliente all’offerta proposta;</li> <li>• <i>O_ACCEPTED</i> = stato finale di successo per l’offerta;</li> <li>• <i>O_CANCELLED / O_DECLINED</i> = stato finale di insuccesso per l’offerta.</li> </ul>
“W_”	<p>Rappresentano gli stati delle attività di lavoro che avvengono durante il processo.</p> <ul style="list-style-type: none"> <li>• <i>W_Afhandelen leads</i> = ricerca di ulteriori informazioni a seguito di incomplete comunicazioni iniziali;</li> <li>• <i>W_Completeren aanvraag</i> = completamento della pre-accettazione della richiesta;</li> <li>• <i>W_Nabellen offertes</i> = classificazione dei clienti per la trasmissione dell’offerta;</li> <li>• <i>W_Validatoren aanvraag</i> = valutazione della risposta all’offerta;</li> </ul>

- 
- |  |   |
|--|---|
|  | <ul style="list-style-type: none"> <li>• <i>W_Nabellen incomplete dossiers</i> = ricerca di ulteriori informazioni;</li> <li>• <i>W_Beoordeelen fraude</i> = indagine e valutazione su sospetto caso di frode;</li> <li>• <i>W_Wijzigen contractgegevens</i> = modifica dettagli del contratto dell'offerta.</li> </ul> |
|--|---|

Inoltre, per ciascun oggetto preso in carica da attività manuali, vi è una distinzione in base al suo ciclo di vita:

- *SCHEDULE*: sta ad indicare che l'oggetto è stato inserito in coda al fine di essere processato in futuro;
- *START*: inizio della lavorazione;
- *COMPLETE*: conclusione del lavoro;

L'attività di *start* risulta essere una sola (che infatti occorre nel 100% dei casi), mentre le attività di *end* sono molteplici e con diverse frequenze di occorrenza. Se ne riporta una tabella riassuntiva prelevata dal tool *ProM*:

Start events			
Total number of classes: 1			
Class	Occurrences (absolute)	Occurrences (relative)	
A_SUBMITTED+complete	13087	100.0%	
End events			
Total number of classes: 13			
Class	Occurrences (absolute)	Occurrences (relative)	
A_DECLINED+complete	3429	26,202%	
W_Validatoren aanvraag+complete	2745	20,975%	
W_Afhandelen leads+complete	2234	17,07%	
W_Completeren aanvraag+complete	1939	14,816%	
W_Nabellen offertes+complete	1269	9,849%	
A_CANCELLED+complete	655	5,005%	
W_Nabellen incomplete dossiers+complete	452	3,454%	
O_CANCELLED+complete	279	2,132%	
W_Beoordeelen fraude+complete	57	0,436%	
W_Wijzigen contractgegevens+schedule	4	0,031%	
W_Validatoren aanvraag+start	2	0,015%	
A_REGISTERED+complete	1	0,008%	
W_Nabellen offertes+start	1	0,008%	

Figura 2: Eventi iniziali e finali (*ProM*)

Inoltre, si riporta di seguito un'immagine descrittiva e significativa per comprendere quali sono le attività più frequenti in base al rapporto *eventi/giorno*:

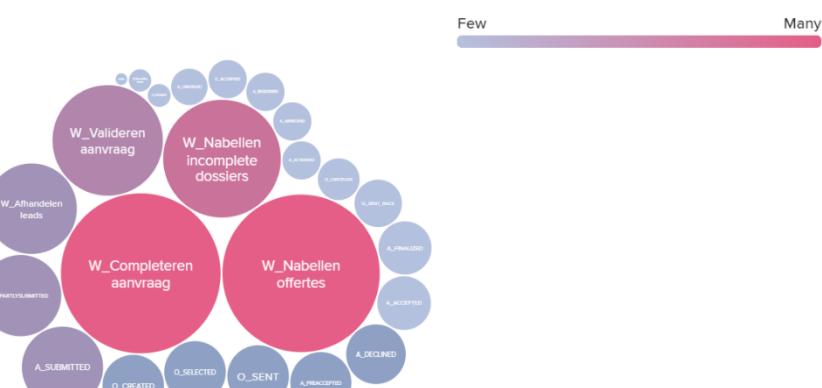


Figura 3: Attività più frequenti (*Celonis*)

## 2.1 Mappa generale del processo

Il processo viene inizializzato da una richiesta di un prestito personale o scoperto attraverso un servizio automatizzato (una pagina web) che si occupa di controllare e verificare l'idoneità della richiesta inoltrata dal cliente. Dopo aver completato la richiesta con informazioni aggiuntive ottenute contattando telefonicamente il cliente, si procede alla verifica dell'idoneità del cliente. Nel caso in cui quest'ultimo risulti idoneo, alla richiesta precedentemente effettuata si integra una proposta offerta ed una nuova valutazione; d'altro canto, se il cliente risultasse non idoneo, la richiesta verrebbe respinta. In quest'ultimo caso, le operazioni che seguono le attività di insuccesso riguardano ad esempio il controllo delle frodi o la mancanza di informazioni essenziali da parte del cliente. Utilizzando il tool *Disco* e visualizzando il 35% delle attività e il 4% dei *path*, è possibile ottenere la seguente mappa generale che riassume quanto detto sopra:

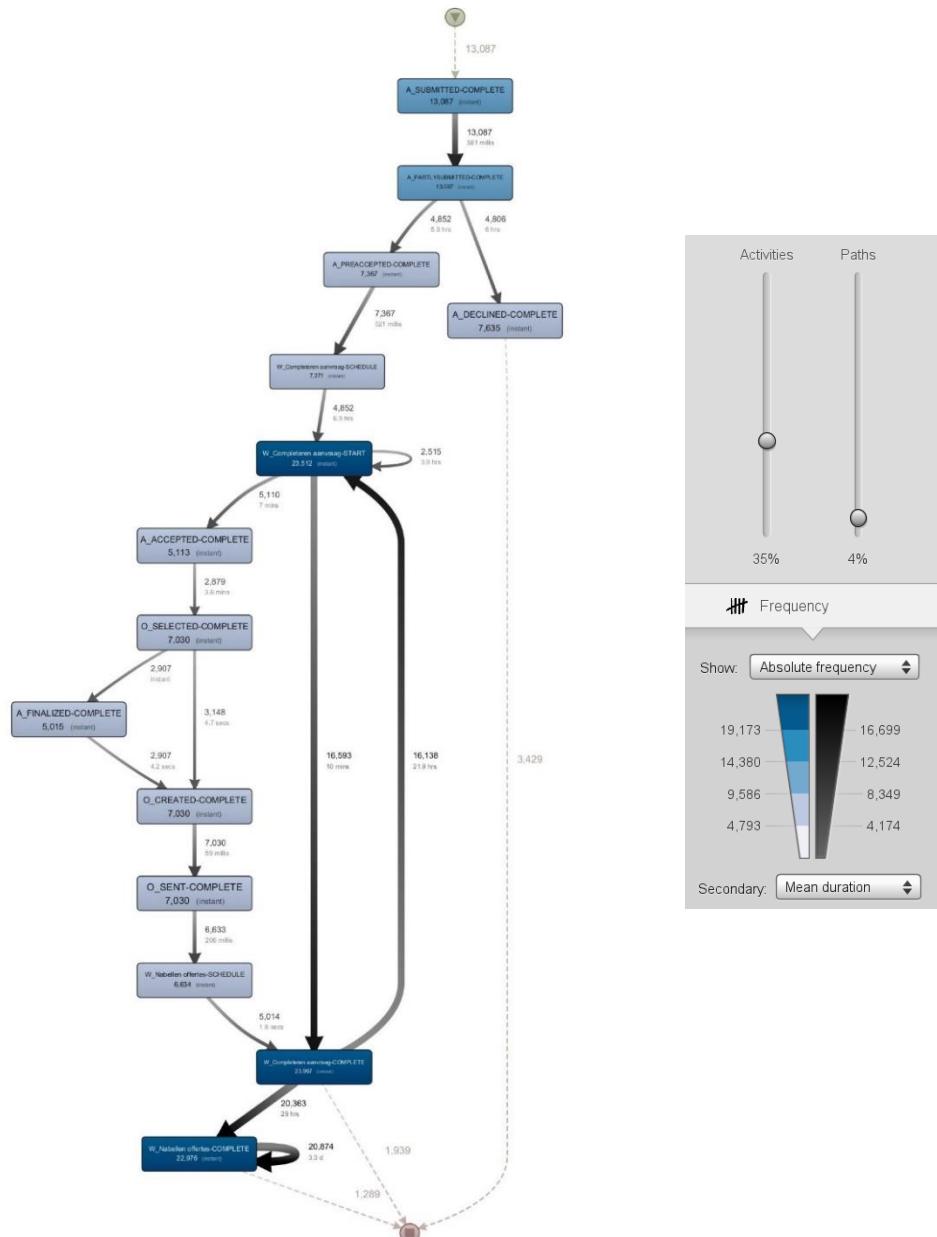


Figura 4: Mappa generale del processo (*Disco*)

In particolare, nel caso del sotto-processo A, esplodendo la percentuale di *threshold* al massimo, si ha la seguente mappa:

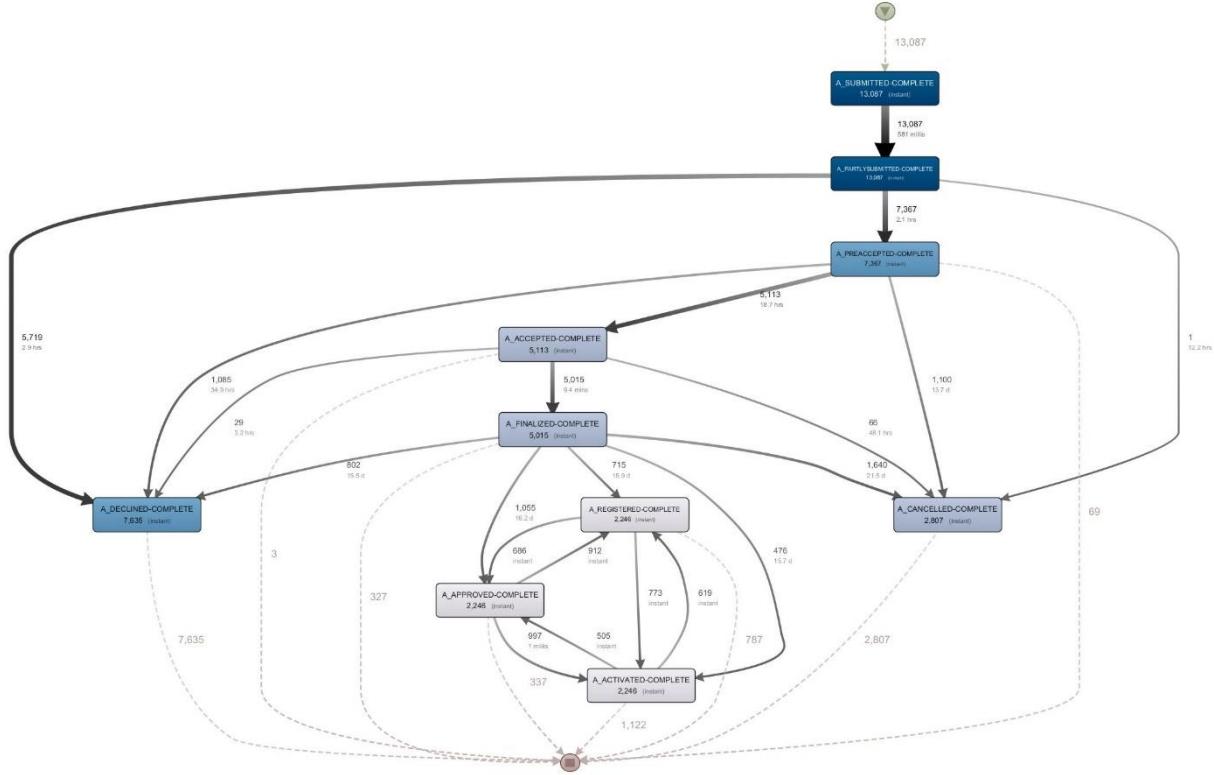


Figura 5: Mappa esplosa del sotto-processo "A" (Disco)

Tale sottoprocesso, coinvolge il 23% degli eventi e il 100% dei case (in quanto l'evento di start è appunto *A\_SUBMITTED*).

Prendendo in esame il sotto-processo O, esplodendo la percentuale di *threshold* al massimo, si ha la seguente mappa, che coinvolge invece il 38% dei case e l'11% degli eventi:

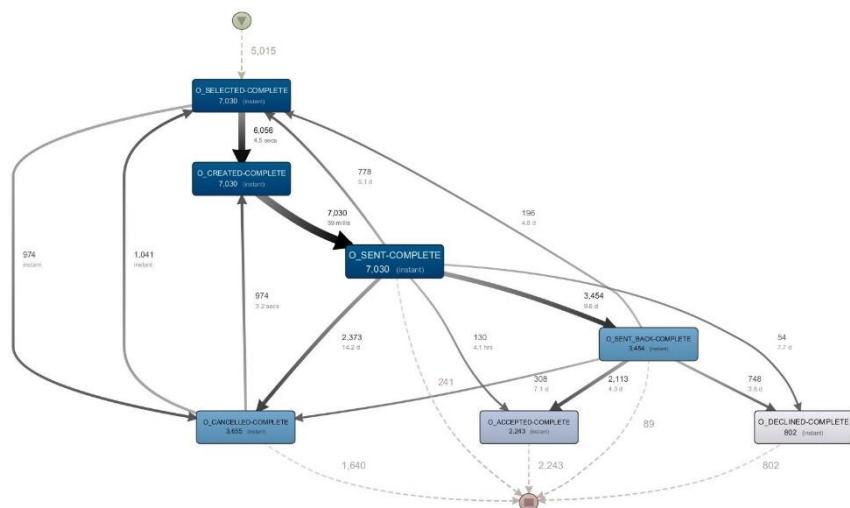


Figura 6: Mappa esplosa del sotto-processo "O" (Disco)

Infine, prendendo in esame il sotto-processo W, si ha la seguente mappa esplosa:

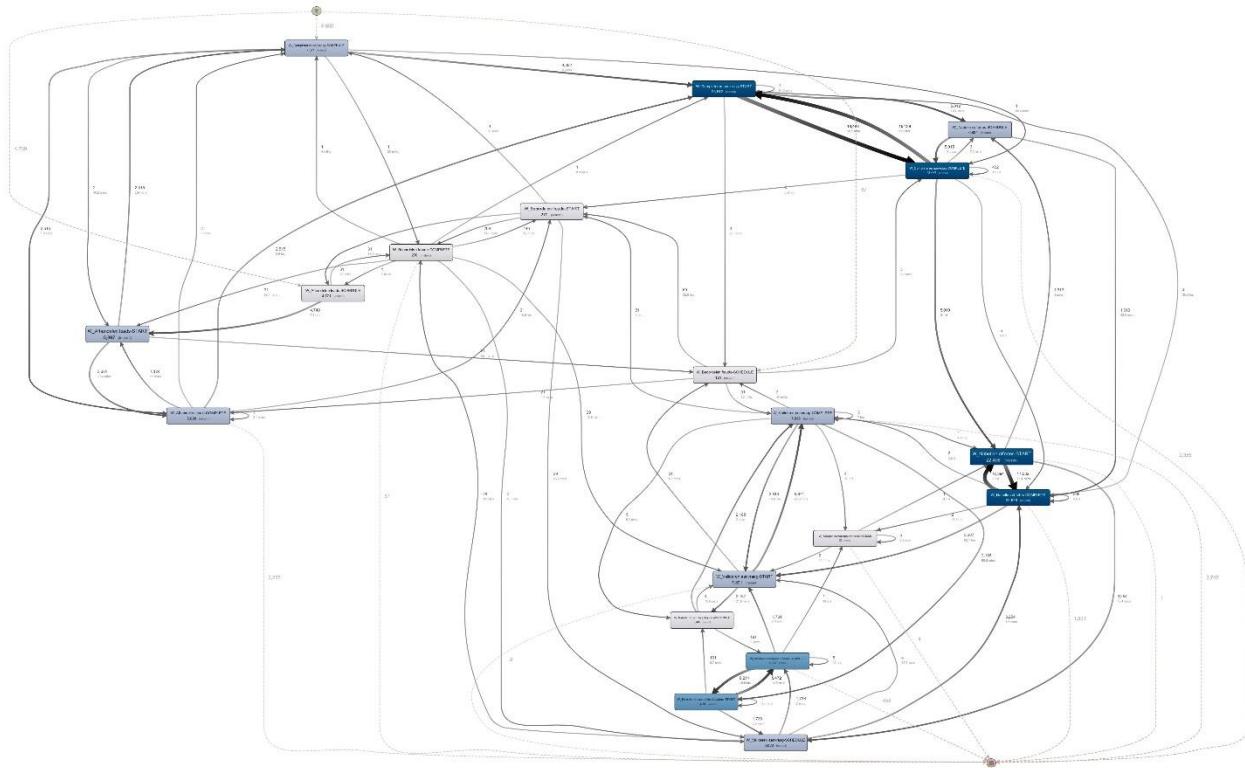


Figura 7: Mappa esplosa del sotto-processo "W" (Disco)

Tale sotto-processo, coinvolge il 73% dei case e il 64% degli eventi; ciò che salta all’occhio immediatamente è il numero di archi e attività che lo rendono più articolato rispetto ai due sotto-processi presi precedentemente in esame.

Per quanto riguarda lo studio delle somme richieste di prestito o come scoperto, è stato analizzato l’attributo *AMOUNT\_REQ* riscontrando alcune anomalie. Nel caso specifico, vi erano 17 richieste di prestito dal valore di 1 e 0 euro che ovviamente sono state rifiutate; mentre la somma più richiesta è di 5000 euro che corrisponde al 12,58% di tutte le richieste.



Figura 8: Top 10 richieste di prestito/scoperto (Disco)

Value	Frequency	Relative frequency
0	11	0 %
1	6	0 %
10	12	0 %
12	3	0 %
13	22	0,01 %
25	35	0,01 %
50	6	0 %
70	3	0 %
98	3	0 %
100	44	0,02 %

Figura 9: Top 10 richieste di prestito/scoperto anomale (Disco)

### 3 Analisi statistica degli eventi

Un passo necessario per un'analisi complessiva è indubbiamente lo studio degli eventi, la loro distribuzione nel tempo e la loro composizione.

Come possiamo notare dal grafico estrapolato dalla sezione *Statistics/Overview* del tool *Disco*, la distribuzione degli eventi totali non è costante nel tempo, ma presenta dei picchi: ad esempio nel giorno 23-02-2012 alle ore 9:38 si presenta un rapporto events/hour pari a 399.

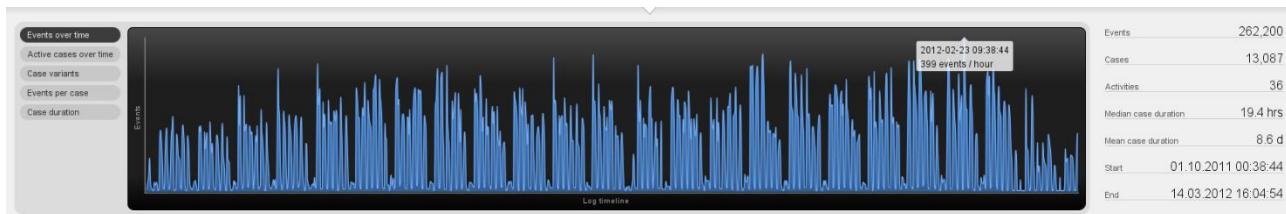


Figura 10: Panoramica statistiche eventi nel tempo (*Disco*)

Come è facilmente osservabile, i giorni in cui vi sono pochi eventi corrispondono generalmente ai giorni festivi e fine settimana, mentre durante le ore di apertura, a differenza del resto della giornata, si verificano dei picchi in termini di numeri di eventi per ora.

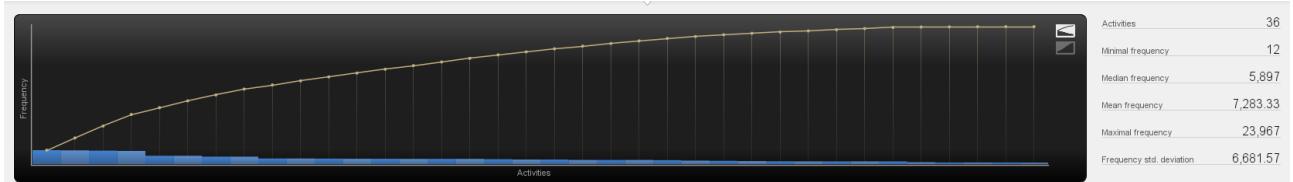
La durata media di un case è di 8.6 giorni, come vedremo in seguito però questa stima è influenzata da alcuni casi particolari che richiedono tempi o troppo brevi (1,855 secondi) o eccessivamente lunghi (137 giorni e 5 ore).

L'andamento dei case attivi a partire dal 01-10-2011 aumenta coerentemente da come ci si potrebbe aspettare da un servizio del genere, all'inizio il numero di casi attivi è molto basso, man mano che passano i giorni questi vanno accumulatosi, l'unico caso di inversione di tendenza è il periodo natalizio, in cui si può ipotizzare una chiusura degli uffici e quindi la presa in carico di sole domande online.



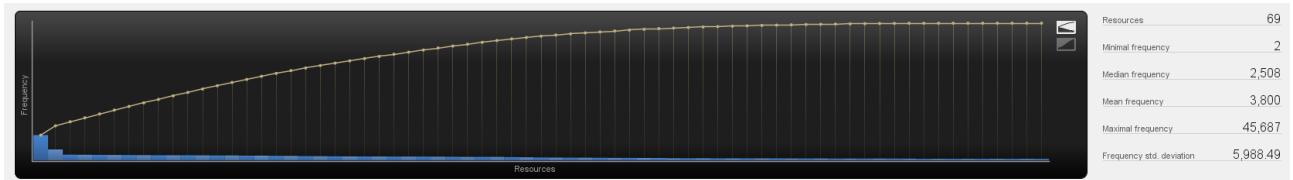
Figura 11: Panoramica statistiche dei case nel tempo (*Disco*)

Preso in esame il grafico dei task (*Figura 10*) si può notare che le 7 attività più frequenti su 36 prendono in carico il 50% degli eventi, nello specifico sono 4 le attività di tipo manuale più frequenti e che coprono il 35.4% degli eventi.



*Figura 12: Panoramica frequenze attività (Disco)*

Le risorse sono 69 e sono identificate da numeri interi, il 17,42% degli eventi utilizza la risorsa 112 che infatti ha la più alta frequenza relativa ed è presente in tutti i case, ma solamente nelle transizioni schedule e complete.



*Figura 13: Panoramica frequenza risorse (Disco)*

Sono inoltre stati analizzati i sotto-processi “A”, “O” e “W”: in particolare, nel caso del sotto-processo “A” si possono notare un numero di eventi che, al contrario dei sottoprocessi “O” e “W”, non sono trascurabili nel fine settimana. Le richieste di prestito, contrassegnate dal nome dell'attività *A\_SUBMITTED*, possono infatti avvenire in qualsiasi giorno, a prescindere che sia feriale o festivo.

### 3.1 Major behaviour e varianti

Rispetto al *major behaviour* le varianti sono 4.366, considerando le prime 10 varianti queste coinvolgono il 49% dei case totali. L'86% delle varianti totali sono coinvolte in un singolo case. Di seguito viene mostrato l'*happy path* e le prime dieci varianti:

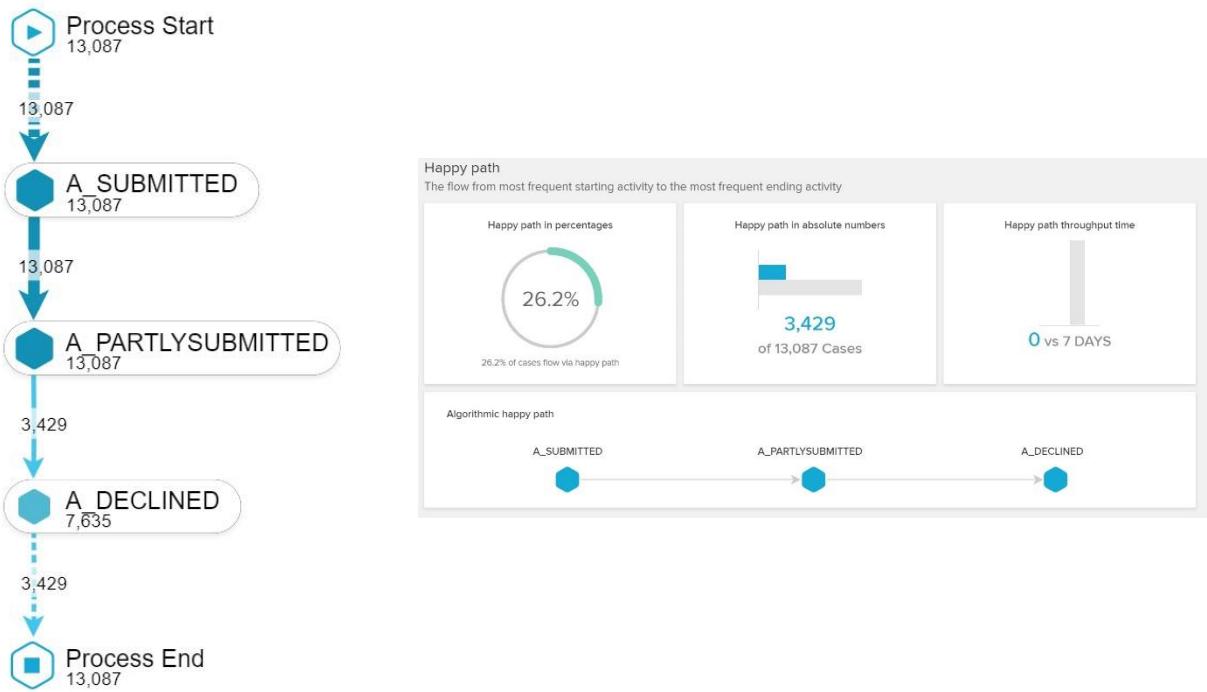


Figura 14: Happy path (Celonis)

Cases (13087)   Variants (4366)				
Variant	Cases	Events	Median duration	Mean duration
Variant 1	3,429	3	37 secs, 828 millis	38 secs, 524 millis
Variant 2	1,872	6	50 mins, 21 secs	5 hours, 10 mins
Variant 3	271	8	2 hours, 5 mins	6 hours, 55 mins
Variant 4	209	10	2 hours, 39 mins	7 hours, 47 mins
Variant 5	160	7	1 hour, 36 mins	6 hours, 17 mins
Variant 6	134	7	2 hours, 8 mins	6 hours, 27 mins
Variant 7	126	12	15 hours, 50 mins	21 hours, 20 mins
Variant 8	93	9	5 hours, 41 mins	13 hours, 24 mins
Variant 9	87	9	10 hours, 37 secs	13 hours, 20 mins
Variant 10	74	14	21 hours, 37 mins	1 day, 6 hours

Figura 15: Top 10 varianti (Disco)

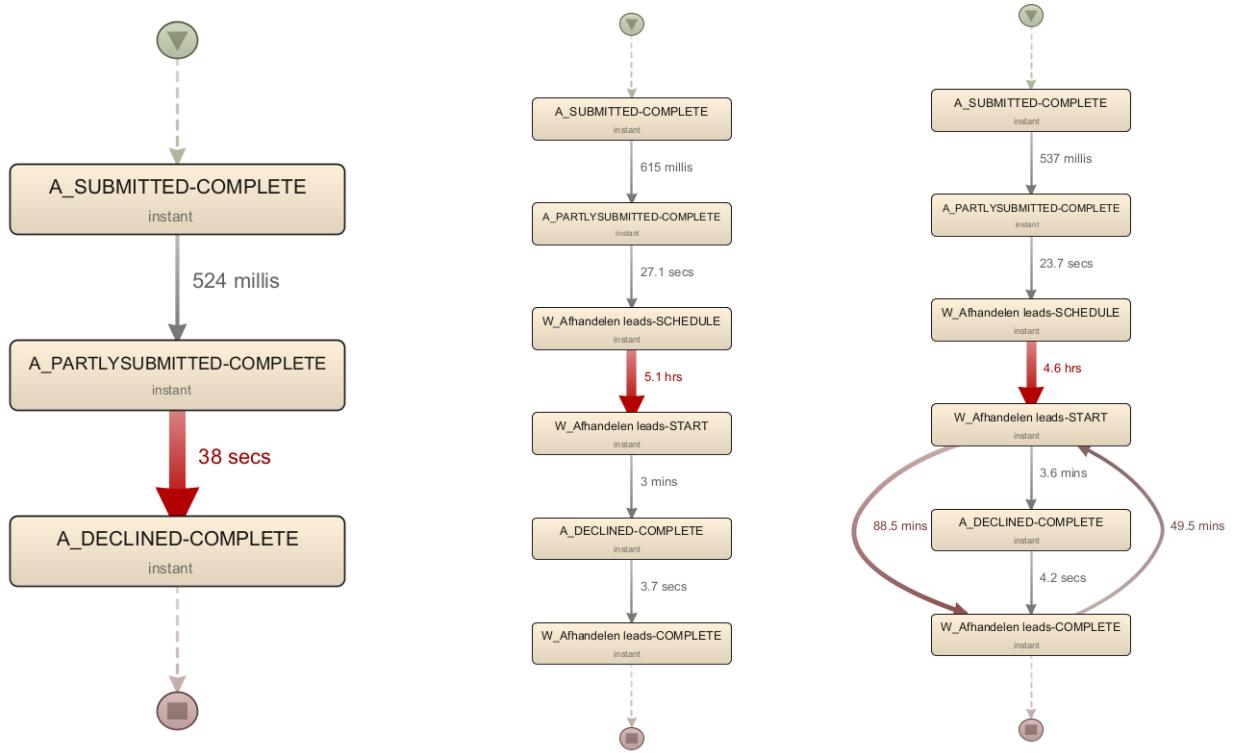


Figura 16: Performance delle prime 3 varianti

Mettendo a confronto le prime tre varianti si nota che condividono le attività iniziali, ma non quelle finali; inoltre, la loro durata media differisce (anche notevolmente) a causa della presenza dell'attività *W\_Afhandelen leads*.

### 3.2 Dotted Chart

Per continuare l'analisi della distribuzione degli eventi nell'arco temporale si può ricorrere all'utilizzo del plugin di ProM *DottedChart*, che attraverso una rappresentazione grafica mette in evidenza l'esecuzione dei vari casi nel lasso temporale che è stato preso in esamina. In particolare, dalla figura sottostante è possibile innanzitutto osservare che sull'asse delle X si ha il *timestamp* mentre su quello delle Y vi sono i vari *case*; nello specifico l'obiettivo di tale analisi era quello di valutare se ci fossero delle possibili fluttuazioni e se in generale si potesse affermare che i vari arrivi dei processi sono costanti per tutto il tempo. Per quanto concerne i diversi arrivi è facilmente visibile che questi sono costanti per tutto il tempo, sono pochi difatti gli eventi che impiegano un tempo maggiore rispetto alla media. Uno in particolare, che viene identificato appunto come *outlier*, è sicuramente il “*W\_Wijzigen contractgegevens*”, tale attività in un caso specifico richiede 102 giorni per essere completato; per questo motivo si è deciso di approfondire ulteriormente su Disco e Celonis l'analisi di questa attività.

Dai risultati ottenuti si può osservare che una risorsa in particolare (la 10912) rallenta il case in cui questo task è coinvolto. Inoltre, si può notare che l'attività viene eseguita solo 4 volte in totale e la sua durata media è di circa 31 giorni, il che mostra come in 2 casi sia rapida la sua esecuzione mentre in altri 2 vi sia un vero e proprio rallentamento. Aldilà di questa particolare attività è possibile stabilire che i casi che richiedono più di 90 giorni per la loro esecuzione sono solo 2, mentre quelli che ne richiedono più di 80 sono 6; si può stabilire comunque che i processi tipicamente richiedono un tempo costante.

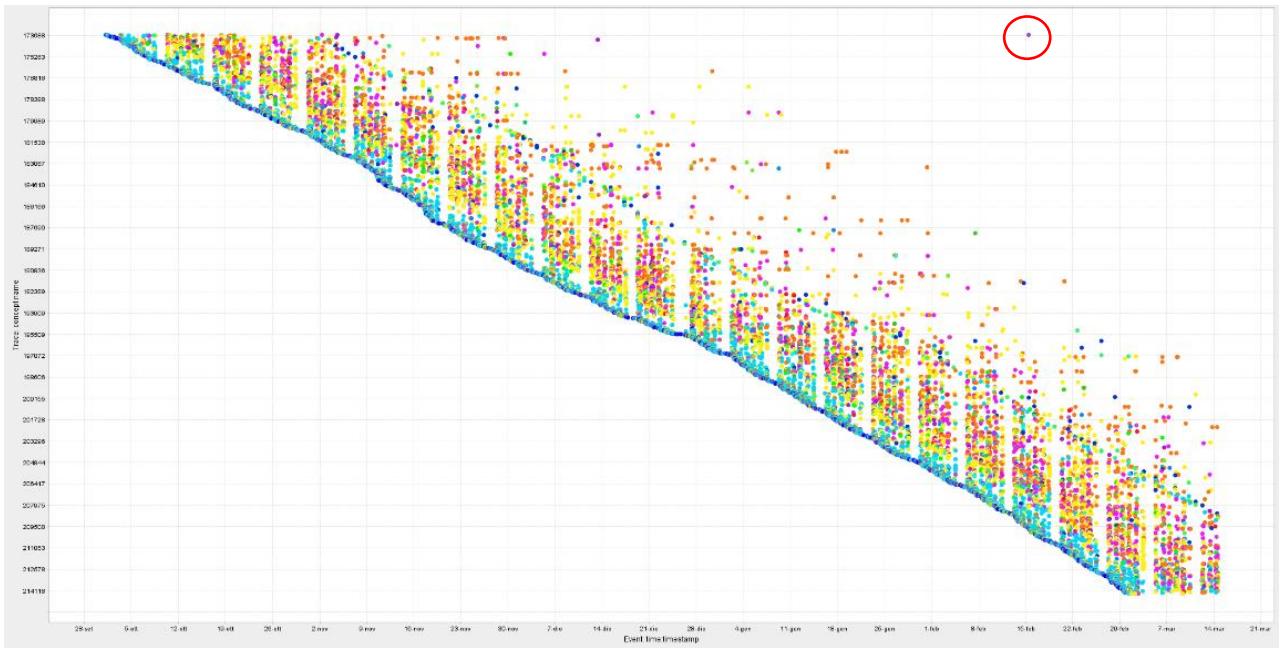


Figura 17: Dotted chart case/timestamp (ProM)

Inoltre, impostando come asse delle ordinate gli eventi per ogni attività, si può notare la loro distribuzione nel tempo:



Figura 18: Dotted Chart eventi/timestamp (ProM)

In particolare, 2 attività sono eseguite in modo costante nel log (*A\_SUBMITTED* e *A\_PARTLYSUBMITTED* sono le attività sempre eseguite in ogni caso), mentre vi sono altre attività che sono eseguite più sporadicamente, come ad esempio *W\_Wijzigen contractgegevens*, che si occupa della modifica dettagli del contratto dell'offerta. Anche in questo, la risorsa 10912 è coinvolta. Si può evidenziare inoltre che dal 29-02-2012 diverse attività non vengono eseguite.

## 4 Process Discovery

Lo scopo del *Process Discovery* è prendere in input un event log e produrre in output un modello di processo, gli algoritmi del *Process Discovery* hanno l'obiettivo di produrre un modello rappresentativo che in pratica catturano i comportamenti più frequenti del log.

Esistono diversi algoritmi di estrazione e si basano sulle due principali caratteristiche degli event log: il rumore e l'incompletezza.

Con il termine *rumore* si vogliono evidenziare i comportamenti che sono rari ed eccezionali all'interno dell'event log, mentre per *incompletezza* si intendono quei log che contengono solo una frazione del comportamento mostrato da un processo di business.

Di seguito verranno riportati i vari modelli estratti evidenziando i risultati ottenuti mediante i vari algoritmi e filtri applicati.

Il primo algoritmo che stato utilizzato per estrarre il modello è l'*Alpha Miner*, questo algoritmo è in grado di generare modelli di processo che presentano attività concorrenti, ma non supporta tracce contenenti attività duplicate ed ignora il problema del rumore nei log.

Il modello estratto risulta essere molto difficile da interpretare infatti le principali cause sono:

- un numero molto elevato di archi (*spaghetti like*)
- presenza di alcuni Eventi (nodi) completamente scollegati nella rete
- la rete ottenuta non gode della proprietà *sound*

Dall'analisi del modello sono state riscontrate le limitazioni dell'algoritmo stesso con un risultato poco soddisfacente:

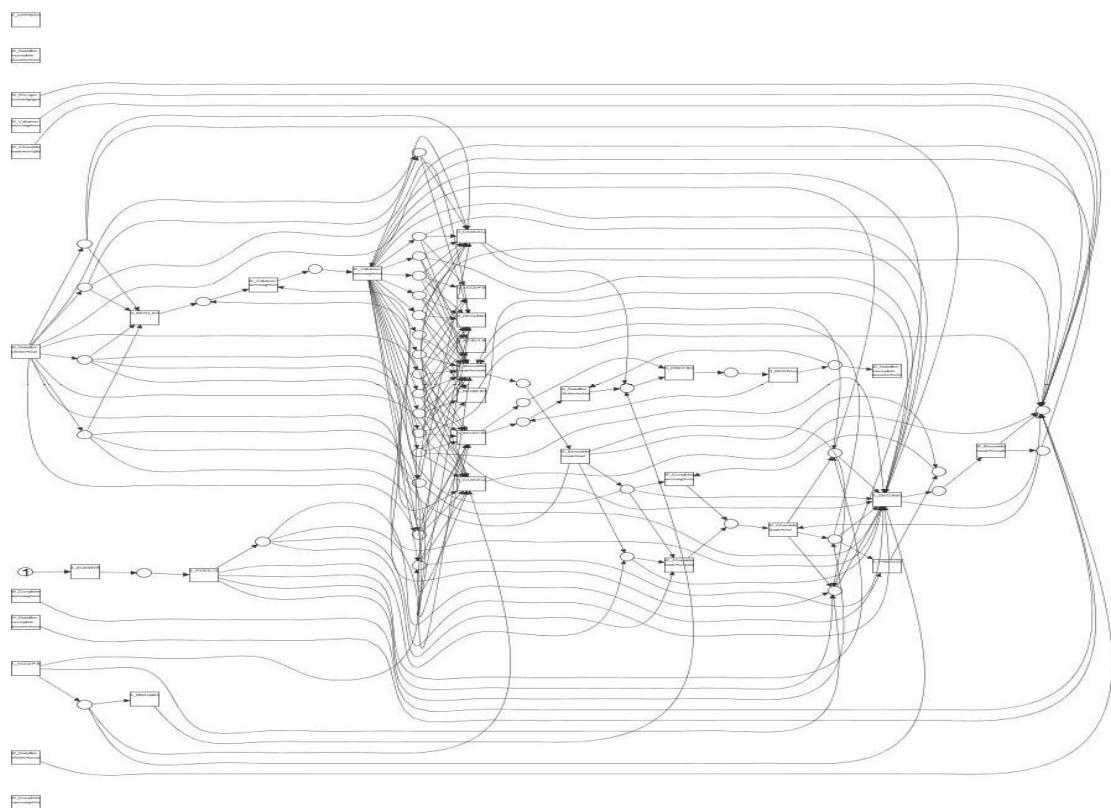


Figura 19: Modello estratto da Alpha miner (ProM)

Infatti, l'algoritmo alpha ha alcune limitazioni quali:

- *presenza di implicit places*
- *loop di lunghezza 1 e 2 non rappresentati*
- *non-local dependencies*
- *representational bias*
- *non tiene conto delle frequenze*

Osservando quindi il risultato dell'*Alpha Miner* si è scelto di applicare il filtro “*filter log using simple heuristics*”, i parametri che sono stati selezionati sono il 100% degli eventi intermedi, mentre come *start* ed *end* sono stati presi quelli complete degli *application*.

Con l'utilizzo di questo filtro è stato possibile risolvere il problema definito “*spaghetti like*”, mentre rimangono i problemi di eventi completamente scollegati dalla rete, oltre al fatto che la rete ancora generata non è sound.

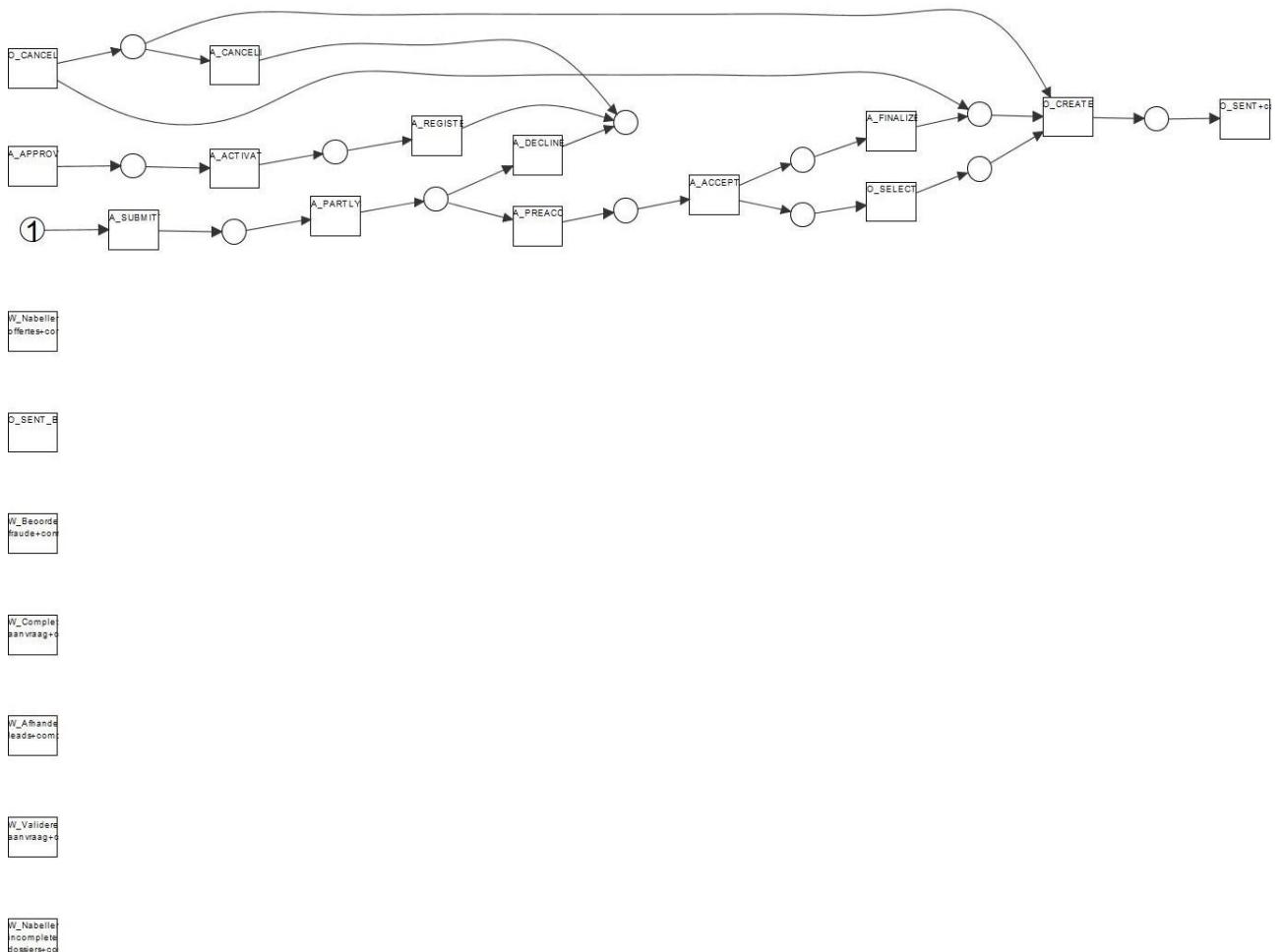


Figura 20: Modello estratto da Alpha miner + "filter log using simple heuristics" (ProM)

Successivamente è stato applicato l'algoritmo *Heuristic Miner*, questo algoritmo rispetto al precedente riesce a superare alcune limitazioni come il filtraggio di componenti eccezionali, quindi con una gestione migliorata del rumore, inoltre permette di derivare la frequenza di esecuzione delle tracce all'interno del log.

Come risultato otteniamo un modello che ha fitness pari al 30,25%, ma che non è una *Petri net* e, per tale motivo, è stata effettuata la “*convert heuristics net into petri net*”, il modello extrapolato dall'algoritmo presenta la connessione di tutti i nodi all'interno della rete però non può definirsi workflow net perché ci sono due end place invece di uno solo.

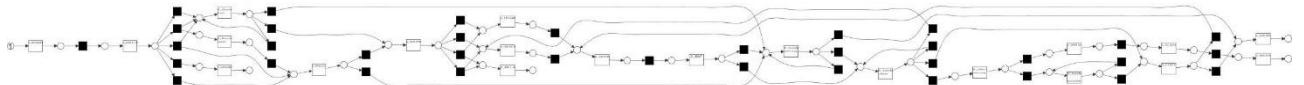


Figura 21: Modello estratto da Heuristic miner + "convert heuristics net into petri net" (ProM)

L'ultima analisi è stata effettuata utilizzando l'algoritmo *Inductive Miner* con *Noise threshold* pari a 0.20 che, rispetto ai modelli precedentemente extrapolati, ci offre risultati molto soddisfacenti. Tutti i nodi della rete sono collegati e il modello gode della proprietà sound, questo perché l'algoritmo riesce ricorsivamente a suddividere l'event log in diversi sotto-log fino a che non contengono singole attività, durante questa suddivisione vengono identificate le principali relazioni tra le diverse attività della rete.

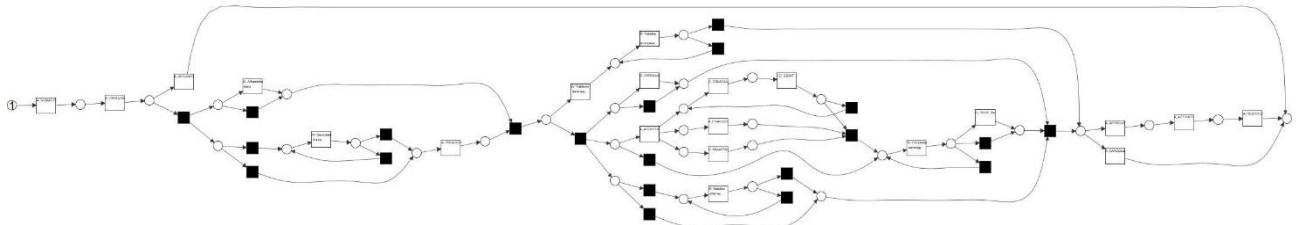


Figura 22: Modello estratto da Inductive miner con Noise threshold al 20% (ProM)

## 5 Process Conformance

L’obiettivo è quello di effettuare un confronto tra il log ed il modello di processo mettendo in evidenza quali tracce contenute nel log deviano dal percorso rappresentato dalla Petri Net ottenuta.

L’analisi della *conformance* è stata eseguita attraverso l’utilizzo dei tool ProM e Celonis. Con ProM è stata valutata la *fitness* che è una delle misure fondamentali per la *conformance*. Il plug-in usato per calcolarla è stato “*Replay a Log on Petri Net for Conformance Analysis*” inserendo come parametri il log filtrato in precedenza e la Petri Net ricavata dall’Inductive Miner. L’analisi è stata effettuata facendo riferimento al miglior modello ottenuto nella fase di process discovery.

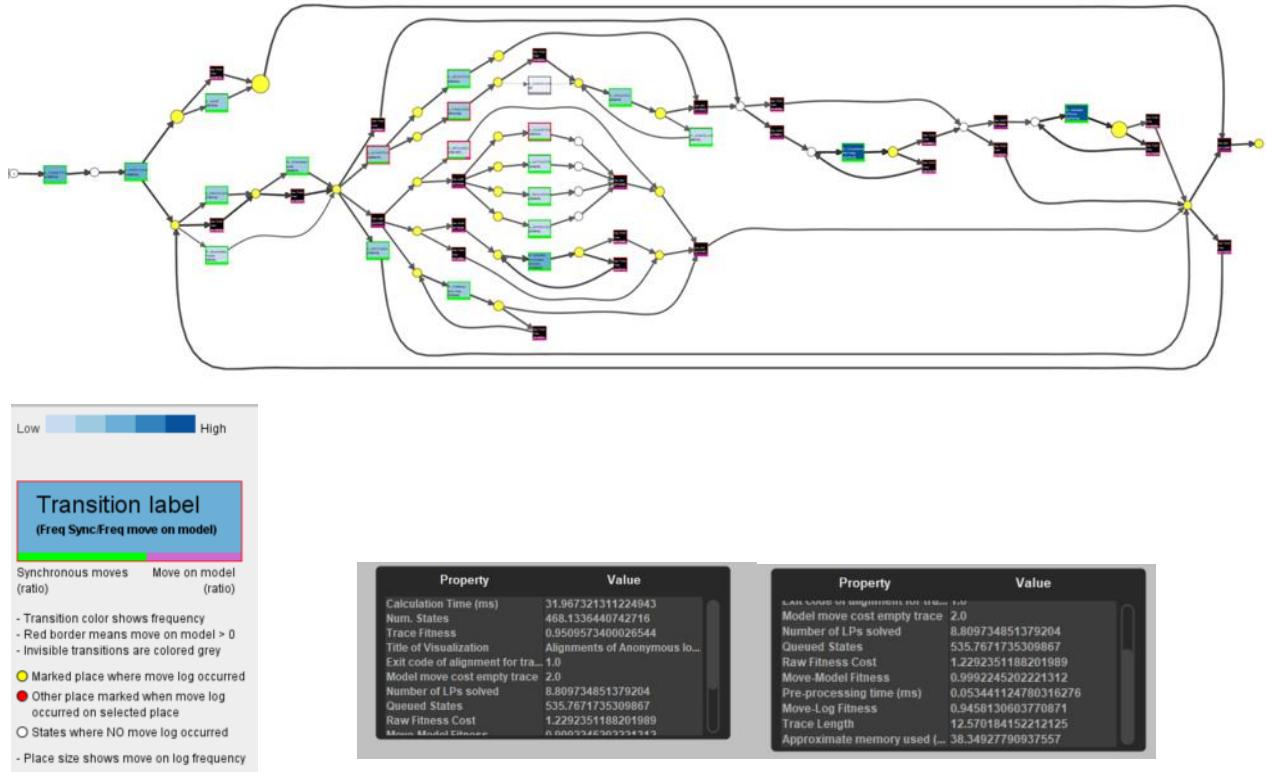


Figura 23: "Replay a Log on Petri Net for Conformance Analysis applicaton" su Petri net ricavata da Inductive miner (ProM)

Osservando la legenda si possono notare i punti in cui il log non corrisponde bene con il modello. Nel modello analizzato sono presenti quattro attività cerchiare in rosso che rappresentano delle violazioni. ProM è stato utilizzato soltanto per valutare la fitness, l’analisi della *conformance* è proseguita su Celonis analizzando le diverse varianti e le relative violazioni.

## 5.1 Conformance sottoprocessi A, O e W

Com'era ipotizzabile, nel caso del sottoprocesso A, i task più frequenti sono “A\_SUBMITTED” e “A\_PARTLYSUBMITTED” che rappresentano le attività iniziali di tutti i case. Si nota inoltre che è presente una sola violazione “A\_DECLINED”, non vi sono ristagni di token all'interno dei place e i restanti task presentano tutti un bordo verde: ciò sta a significare che c'è sincronismo tra le tracce del modello e l'event log.

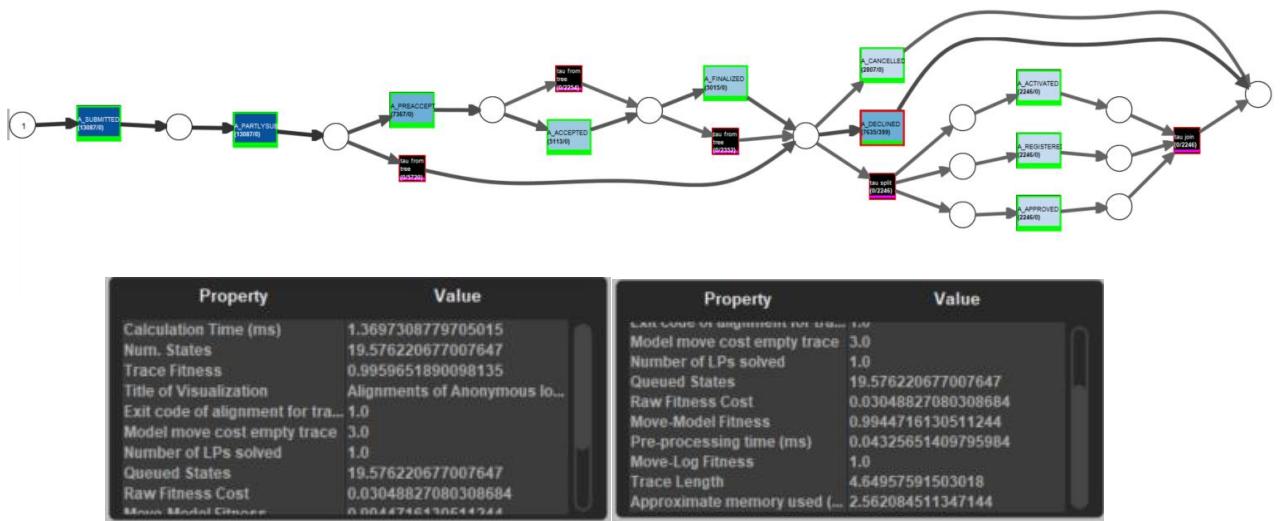


Figura 24: Conformance sublog A (ProM)

Per quanto riguarda i sottoprocessi “O” e “W”, si possono notare alcuni ristagni di token, ma nessuna violazione: come detto in precedenza, infatti i sottoprocessi, presi singolarmente, sono consistenti.

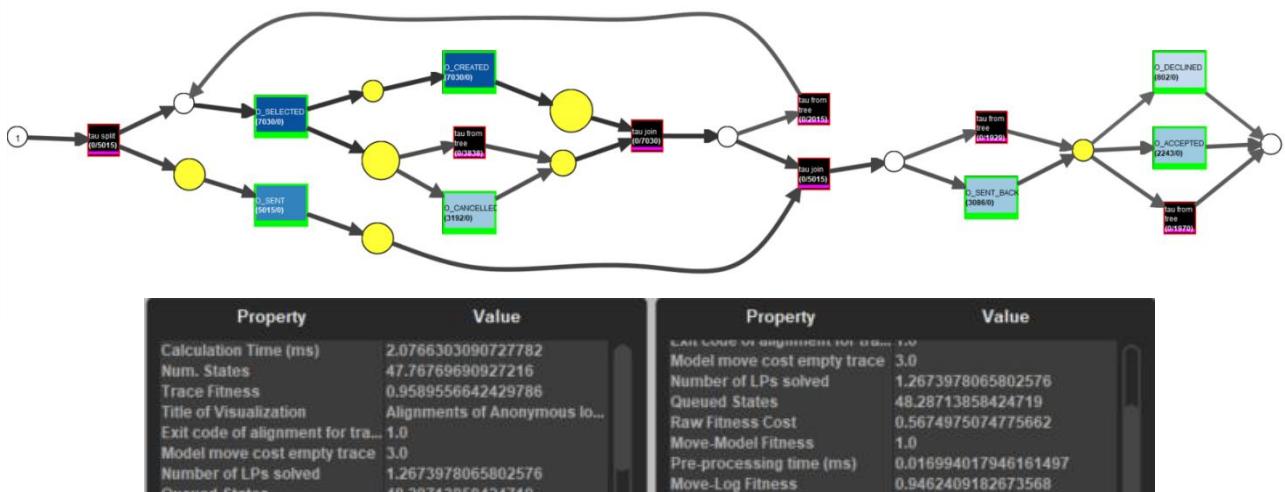


Figura 25: Conformance sublog O (ProM)

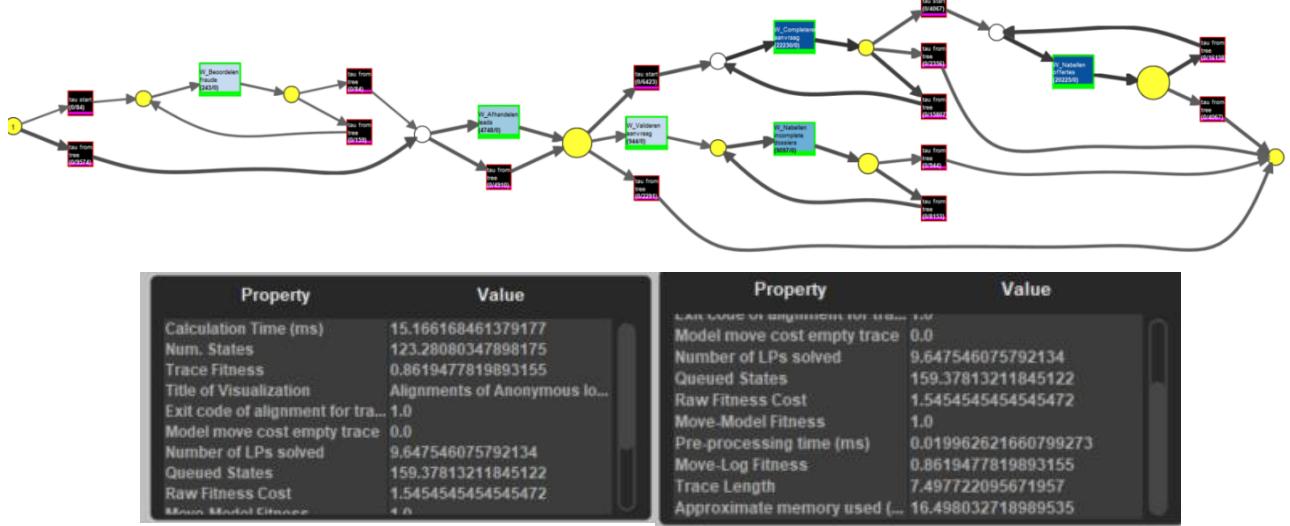


Figura 26: Conformance sublog W (ProM)

## 5.2 Conformance Celonis

Per formare il modello “*to be*” sono state prese le due varianti più comuni (2/1401) che coprono il 43% dei case

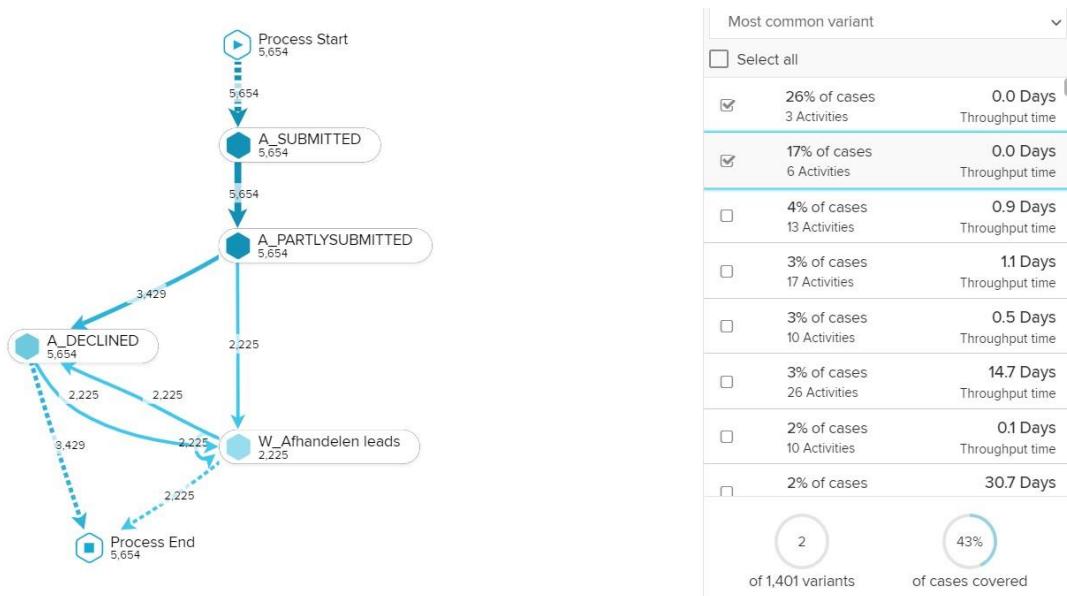


Figura 27: Modello “*to be*” (Celonis)

Confrontando il modello *to be* con il modello *as is* si nota che i case conformi sono 5.65k contro i 7.43k, in totale le violazioni presenti sono 20. Nell'*allowlist* sono state inserite 3 di queste, migliorando la *conformance* del 17%.

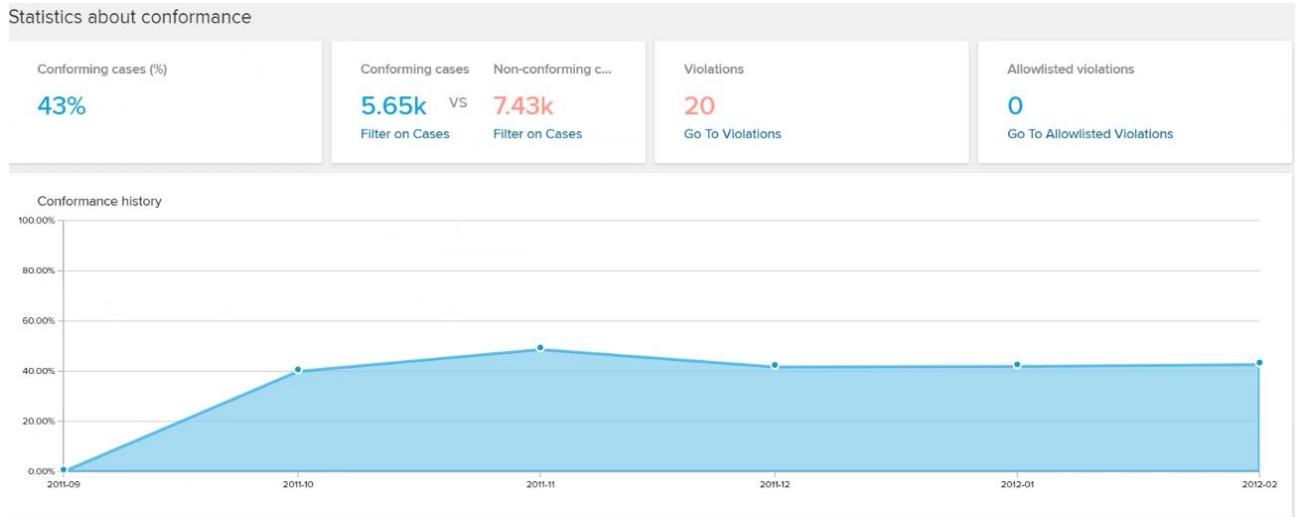


Figura 28: Conformance (Celonis)

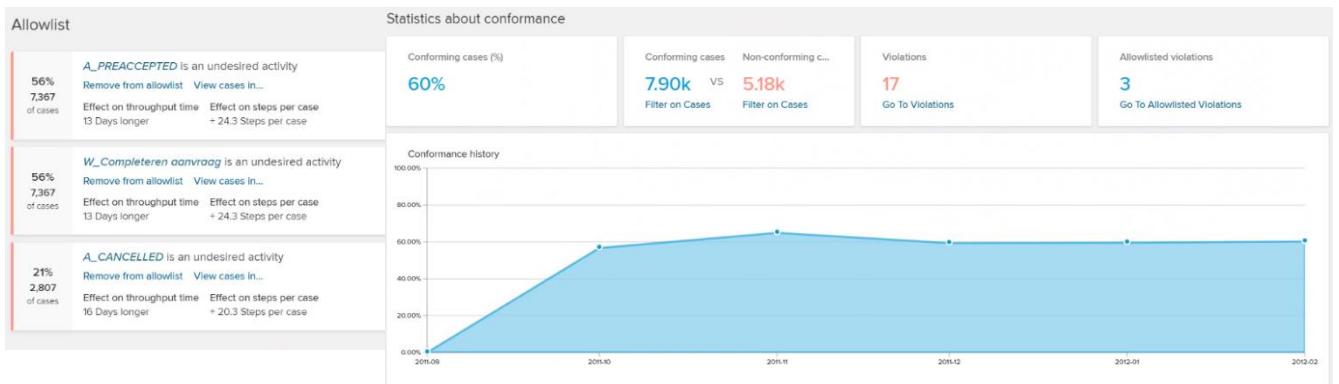


Figura 29: Conformance con modifiche nell'allowlist (Celonis)

### 5.3 Considerazioni sulle varianti

Dal processo preso in analisi, ci si è accorti che il numero delle varianti è estremamente elevato ed inoltre molte di queste rappresentano un solo case.

Le cause ipotizzabili possono essere di due tipologie:

- La possibilità di intrecciare diverse attività (“A”, “O” e “W”) genera numerose combinazioni;
- Le tracce di esecuzione per il controllo dell’idoneità possono variare da cliente a cliente (dati incompleti, controlli, ...);

## 6 Performance Analysis

L'analisi delle performance del flusso di processo è stata effettuata utilizzando Disco, la metrica che è stata presa in considerazione è la durata media, in quanto le altre non davano informazioni rilevanti. Considerando la percentuale di *threshold* pari a quella considerata nella mappa generale del processo, dal modello sottostante si può notare che le attività più onerose in termini di tempo sono:

- *W\_Nabellen\_offertes-COMPLETE* -> *W\_Nabellen\_offertes-COMPLETE* (3,3 giorni)
- *W\_Completeren\_aanvraag\_COMPLETE* -> *W\_Nabellen\_offertes-COMPLETE* (29 ore)
- *W\_Completeren\_aanvraag\_COMPLETE* -> *W\_Completeren\_aanvraag\_START* (21,9 ore)

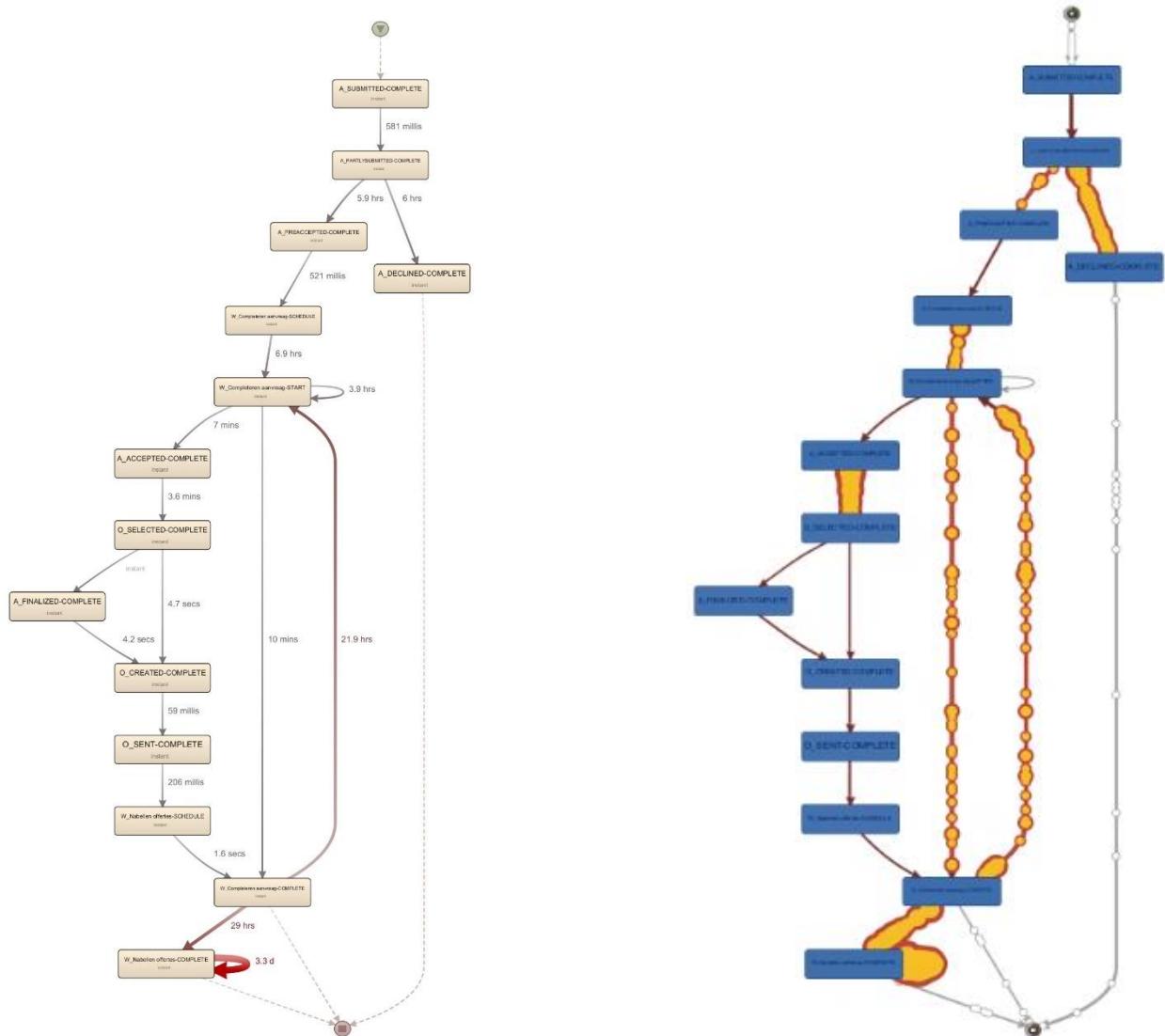


Figura 30: Performance mappa generale (Disco)

## 6.1 Performance sottoprocessi A, O e W

Anche in questo caso, prendendo i sottoprocessi divisi, si perde di generalità; infatti, nella mappa generale compaiono dei colli di bottiglia che invece non sono presenti nelle mappe specifiche e viceversa.

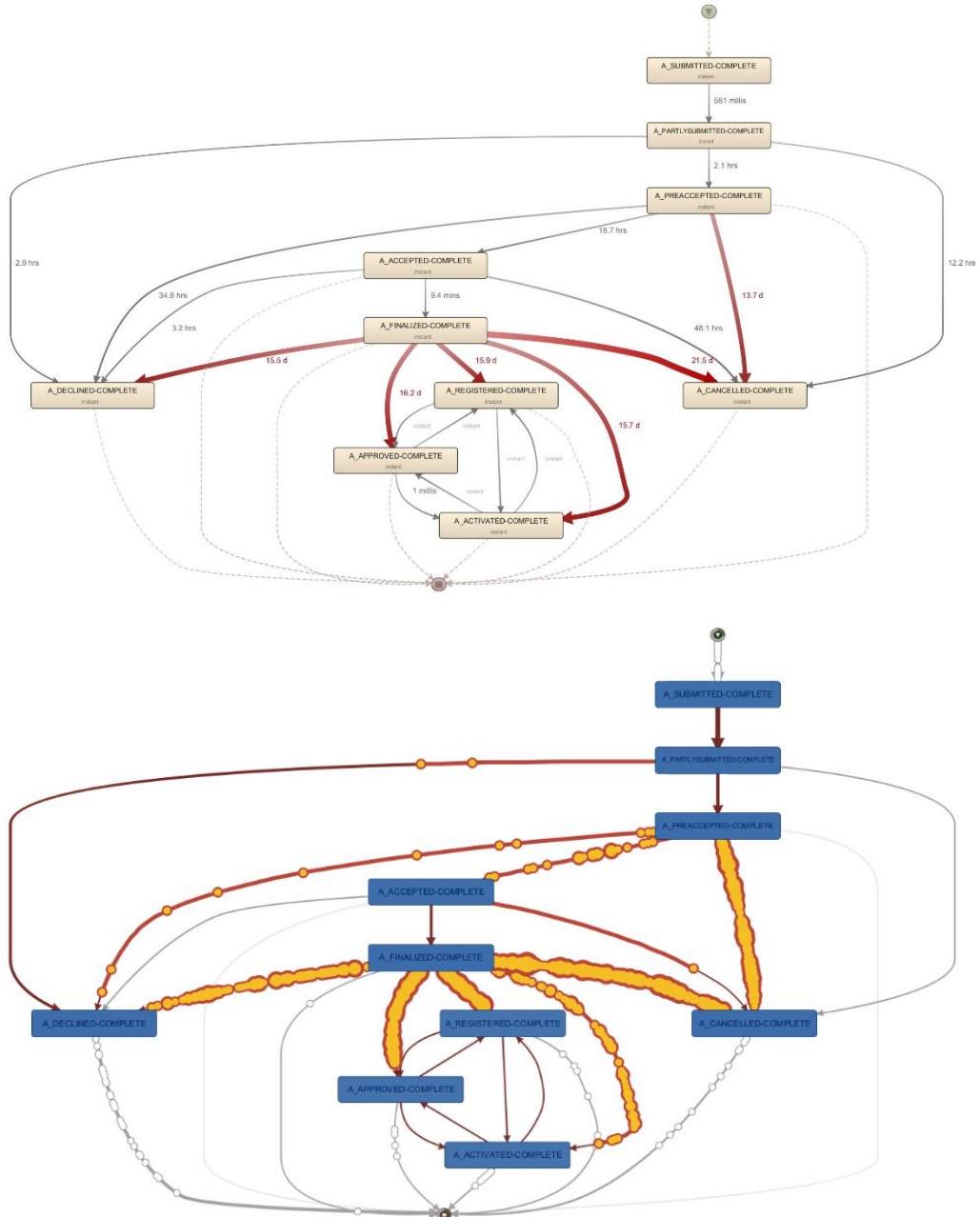


Figura 31: Performance sublog A (Disco)

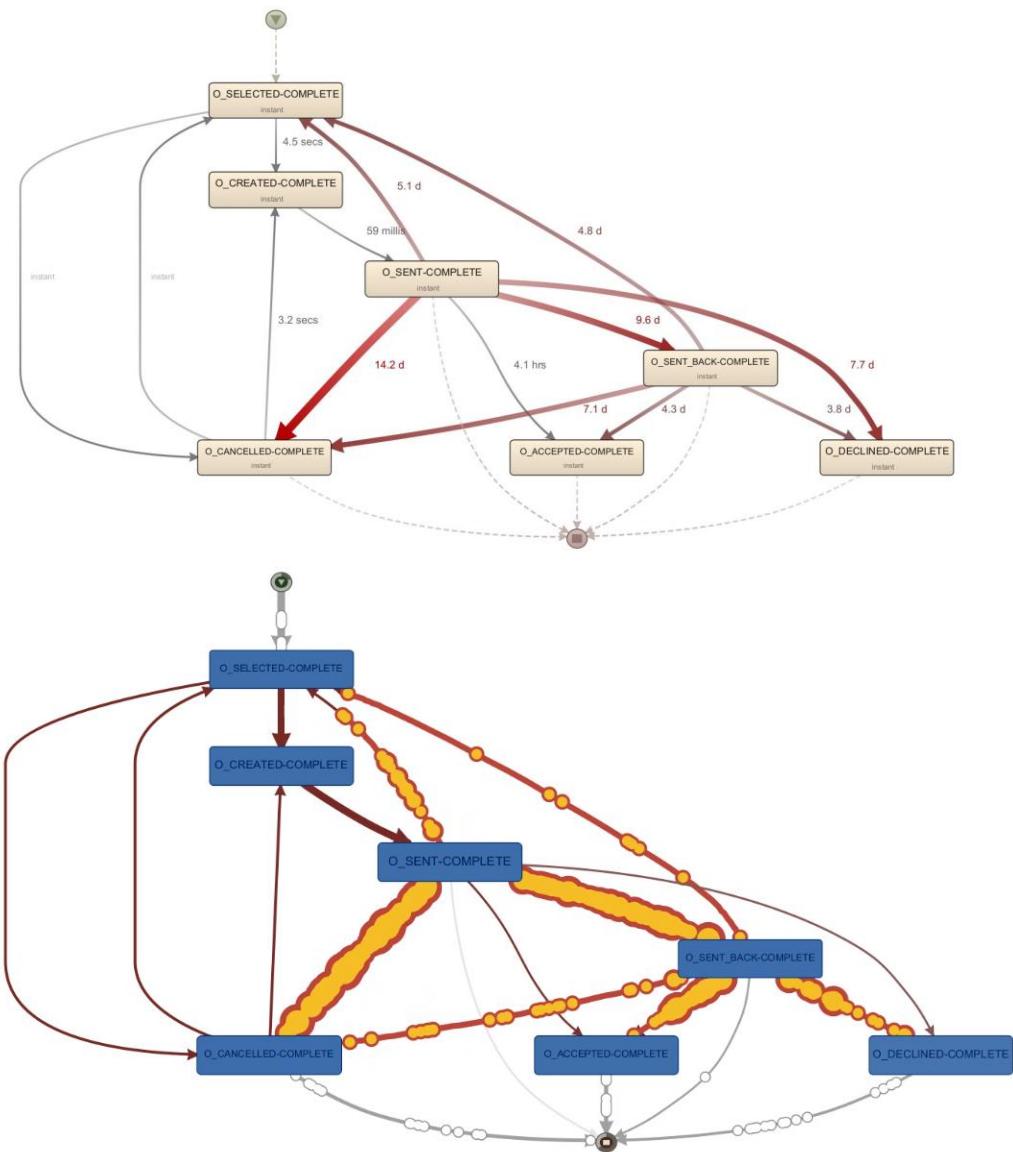
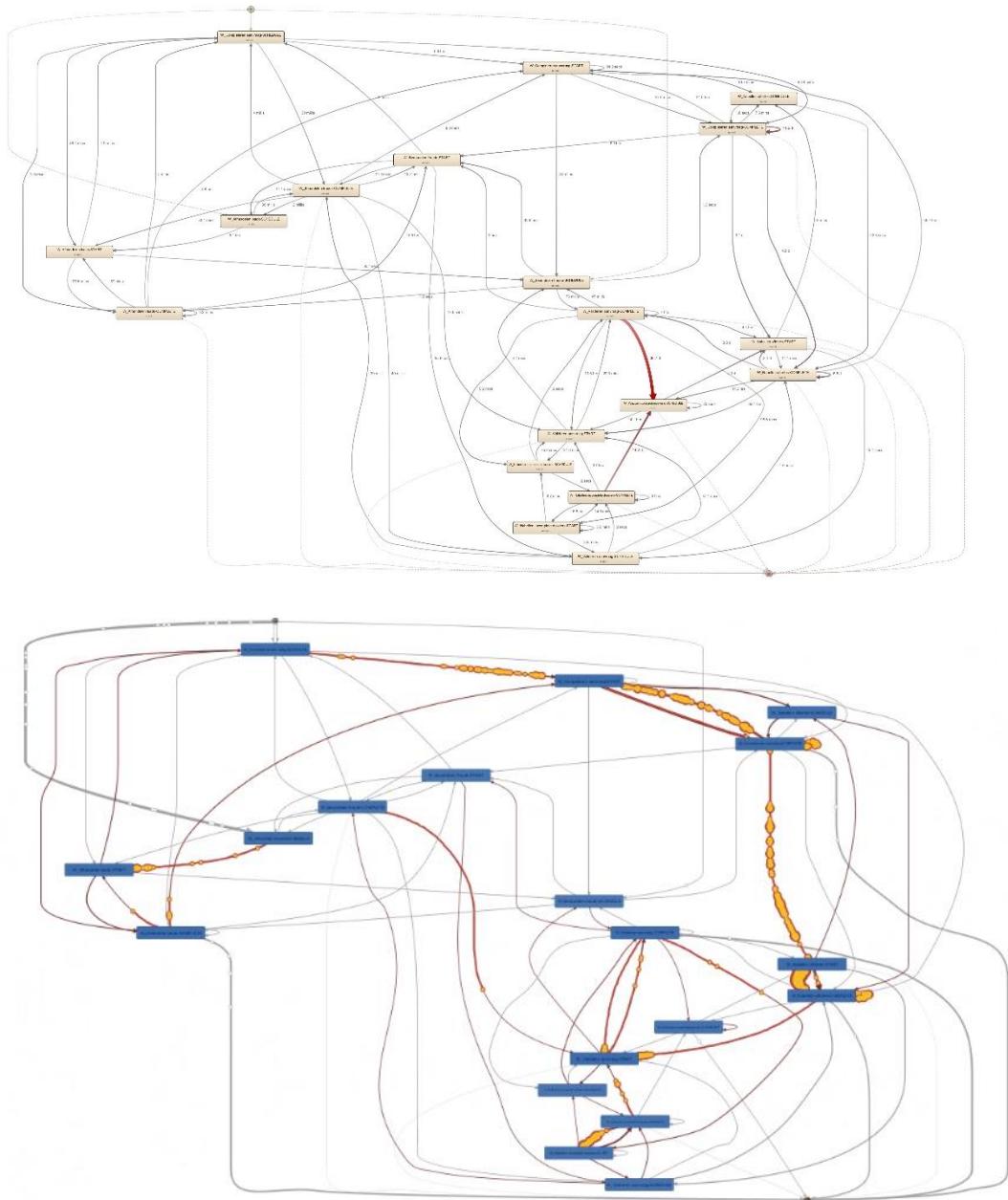


Figura 32: Performance sublog O (Disco)



*Figura 33: Performance sublog  $W$  (Disco)*

## 6.2 Performance Celonis

Sono stati successivamente approfonditi e analizzati i colli di bottiglia su Celonis, questo software è in grado di dare informazioni più precise sui colli di bottiglia rispetto al throughput del processo, considerato l'incidenza di ogni collegamento presente nel processo considerato.

Nella figura successiva vengono inoltre riportati la distribuzione delle tracce, come si può facilmente notare la maggior parte dei case viene eseguito tra 0 e 4 giorni e coinvolge il 56% dei case (7389 su 13087).

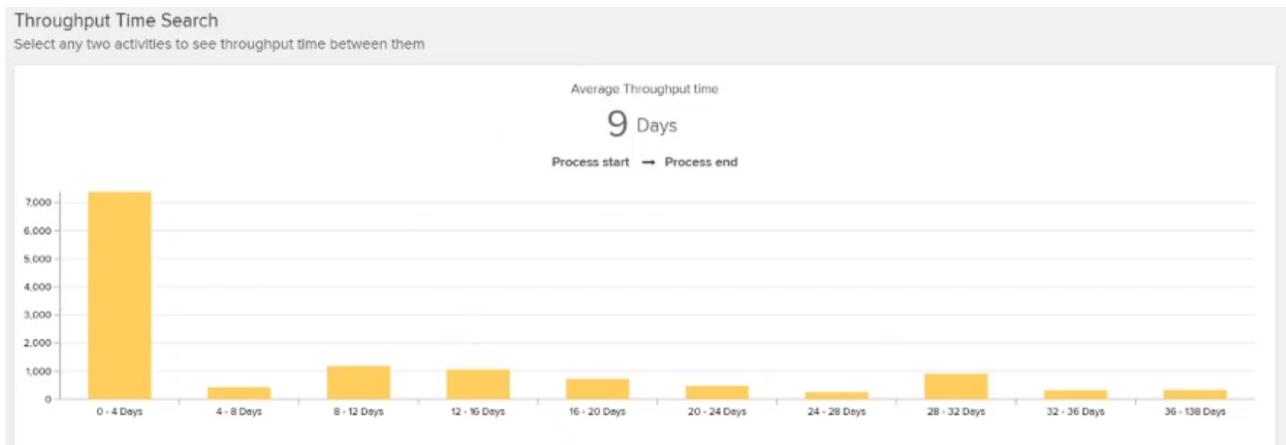


Figura 34: Throughput generale (Celonis)

Il throughput medio è di 9 giorni poiché il 12% dei casi viene eseguito dai 28 giorni in su, fino ad arrivare ad alcuni casi eccezionali che superano gli 80 giorni.

Successivamente si sono andati ad analizzare i tempi sui singoli collegamenti, nella figura successiva sono riportati i Bottlenecks più importanti che incidono in modo significativo all'throughput medio, la sezione sui Bottlenecks è stata effettuata prendendo singolarmente i collegamenti senza combinarli tra di loro.

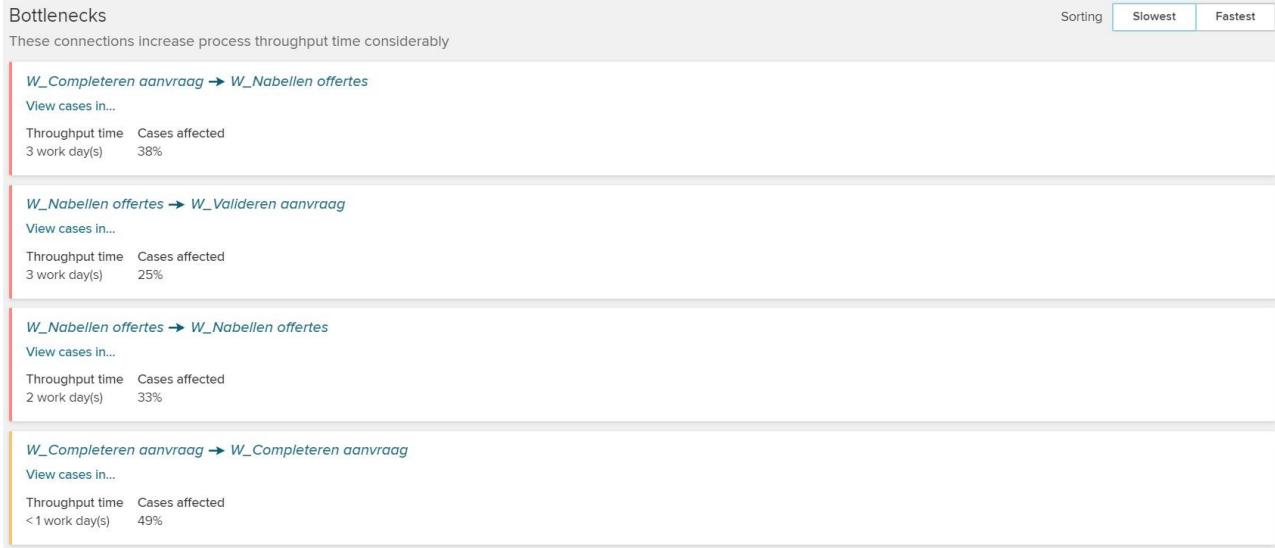


Figura 35: Bottleneck (Celonis)

Il primo collegamento è tra *W\_Completeren aanvraag -> W\_Nabellen offertes* è molto importante come collo di bottiglia, infatti incide di 3 giorni sui 9 totali e coinvolge il 38% dei casi, analizzando quindi questo caso specifico si passa da 9 a 19 giorni, si può facilmente notare che preso un range da 0 a 18 giorni i casi coinvolti sono il 58%, così come indicato nella figura successiva.

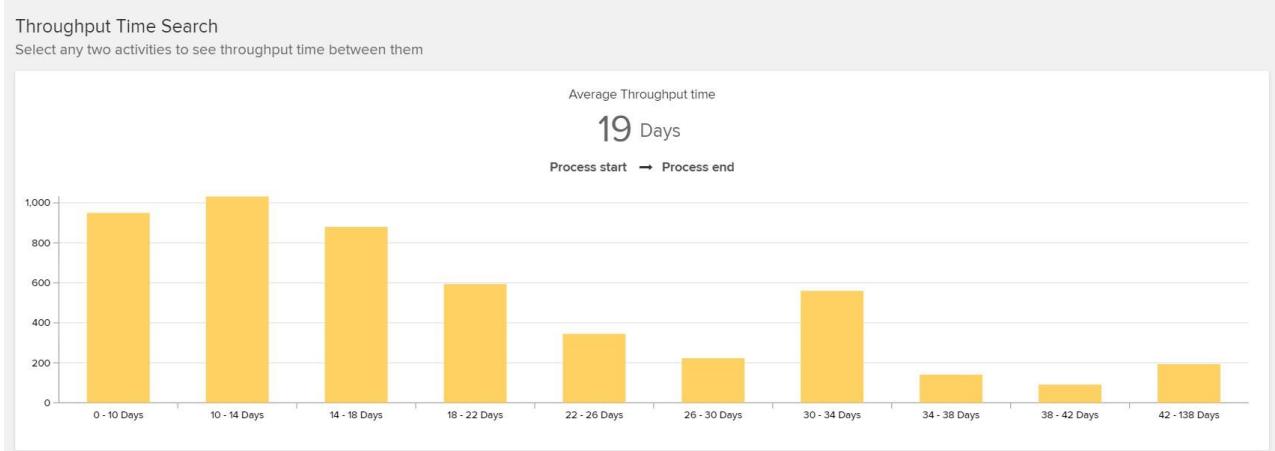


Figura 36: Throughput con "W\_Completeren\_aanvraag -> W\_Nabellen\_offertes" (Celonis)

Anche il secondo collegamento tra *W\_Nabellen offertes ->W\_Validatoren aanvraag* risulta essere rilevante, questo incide anche di 3 giorni e coinvolgere il 25% dei casi, analizzando quindi questo caso si passa da 9 a 17 giorni, considerando il range da 0 a 18 giorni qui i casi coinvolti sono il 63% così come indicato nella figura successiva.

Throughput Time Search  
Select any two activities to see throughput time between them

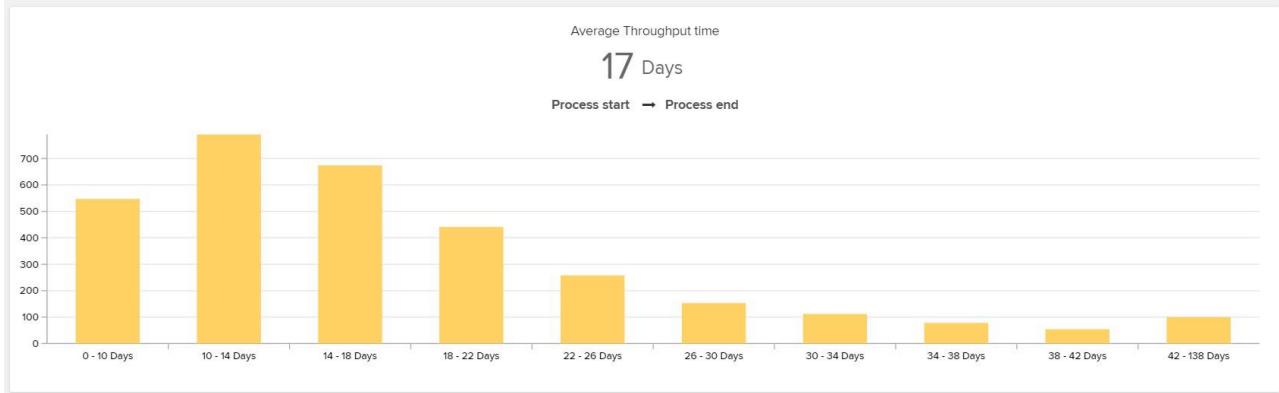


Figura 37: Throughput con “W\_Nabellen offertes -> W\_Valideren aanvraag”

Il terzo è ultimo collegamento è  $W_{Nabellen\ offertes} \rightarrow W_{Nabellen\ offertes}$ , l'incidenza è di 2 giorni e coinvolge il 33% dei casi, il throughput passa da 9 a 20 giorni, considerando sempre il range da 0 a 18 giorni il numero di casi coinvolti è del 55%, così come indicato nella figura successiva.

Throughput Time Search  
Select any two activities to see throughput time between them

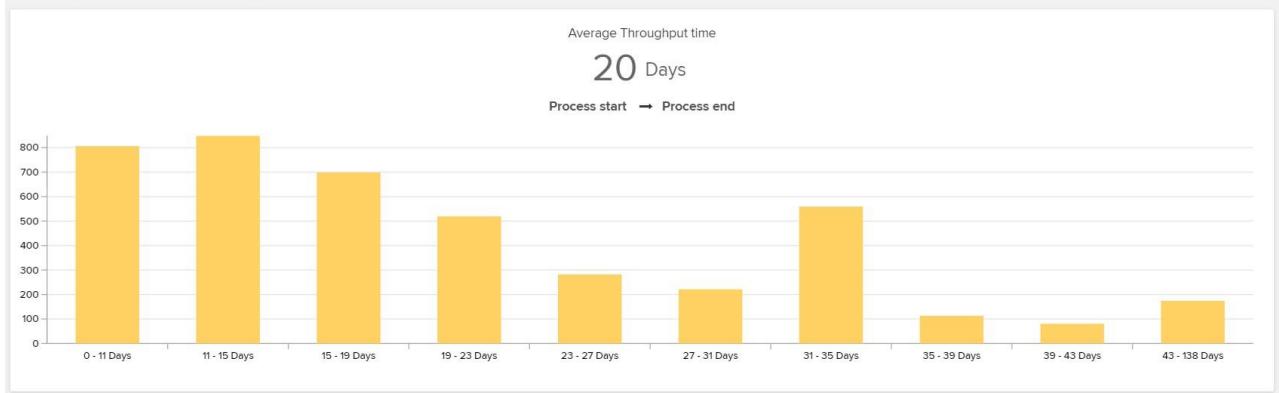


Figura 38: Throughput con “W\_Nabellen offertes -> W\_Nabellen offertes”

## 7 Resource analysis

Dopo aver focalizzato l'attenzione, nei capitoli precedenti, sulle tecniche di *Process Discovery* e *Conformance checking*, lo studio viene completato in questa sezione con l'analisi delle risorse e delle social network.

Qui di seguito è stata riportata la sezione di Disco inerente alle risorse, queste sono pari 69, di cui solo una è interessata all'evento di *start* mentre 54 risorse partecipano almeno una volta all'evento di *end*.

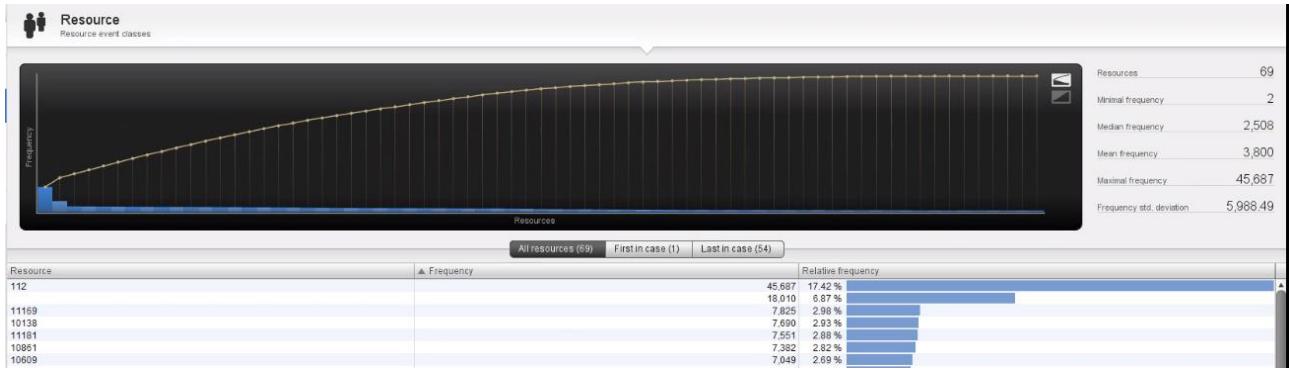


Figura 39: Panoramica frequenza delle risorse (Disco)

### 7.1 Mine for Handover-of-Work Social Network

Utilizzando un altro software per l'analisi delle risorse, si è deciso di sfruttare alcuni plugin di ProM, il primo è il *Mine for Handover-of-Work Social Network*. Questa tecnica mette in risalto il rapporto tra le risorse, ovvero chi passa il lavoro a chi, questo è espresso tramite un rapporto di unidirezionalità.

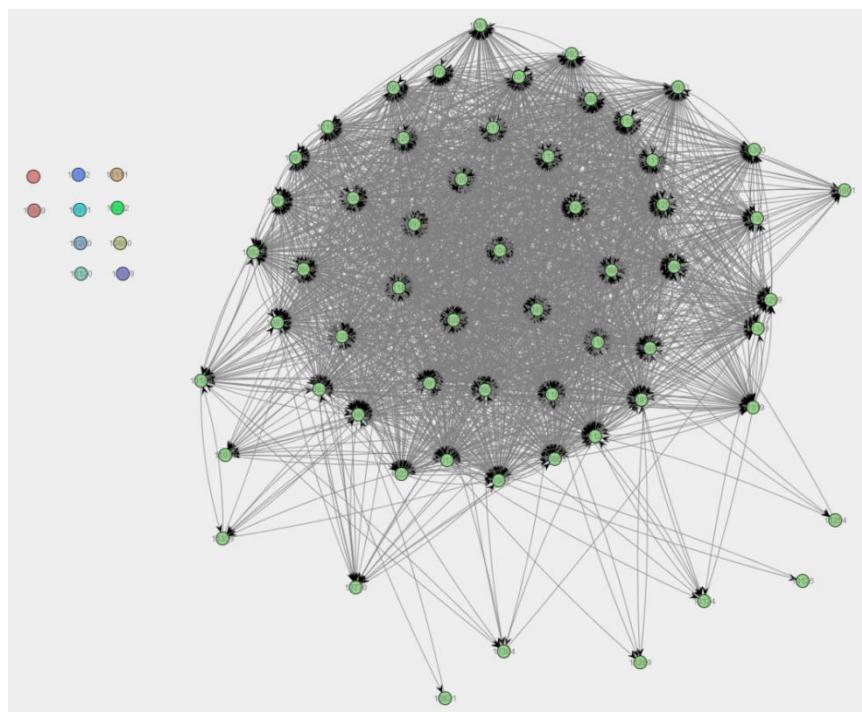


Figura 40: Handover of Work (ProM)

Come parametro di ranking è stato selezionato il *degree*, come ben visibile dall'immagine esistono poche risorse scollegate dall'insieme:

- Quella senza nome
- 10 862
- 11 111
- 11 029
- 10 971
- 11 202
- 11 200
- 10 880
- 11 120
- 10 859

Inoltre, si può notare che le risorse più coinvolte sono rappresentate più vicine al centro del grafico, mentre quelle più distanti come ad esempio 10821 risultano avere meno scambi di lavoro con le altre.

Spuntando la casella su “skretch by degree ratio”, si ottiene la seguente figura:

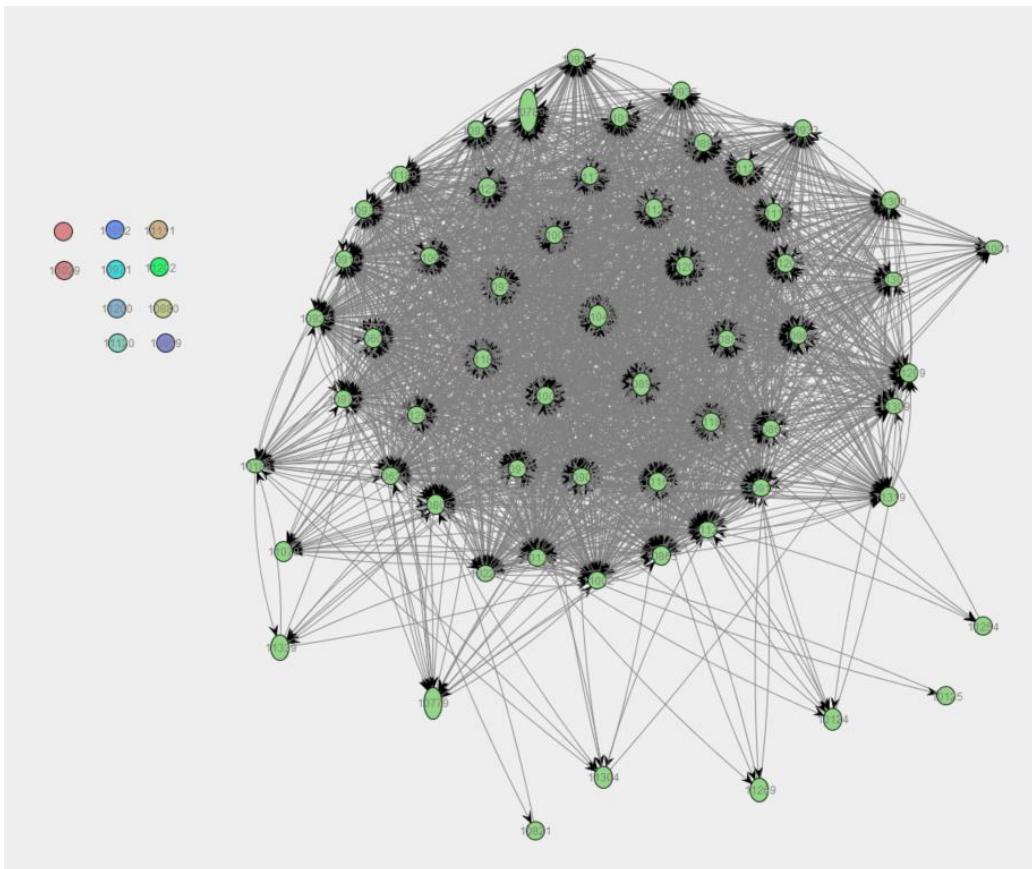


Figura 41: Handover of Work + skretch by degree ratio (Prom)

Si può notare come alcune attività abbiano una forma ellittica coi lati schiacciati: questo sta a significare che gli archi entranti sono maggiori degli archi uscenti. E ciò riguarda in particolare le attività:

- 10 789
- 10 779

Mentre altre attività al contrario hanno pochi archi entranti e molti archi uscenti:

- 11 001
- 10 188

Quanto detto è rimarcato dalla seguente figura (spuntando la casella “size by ranking”):

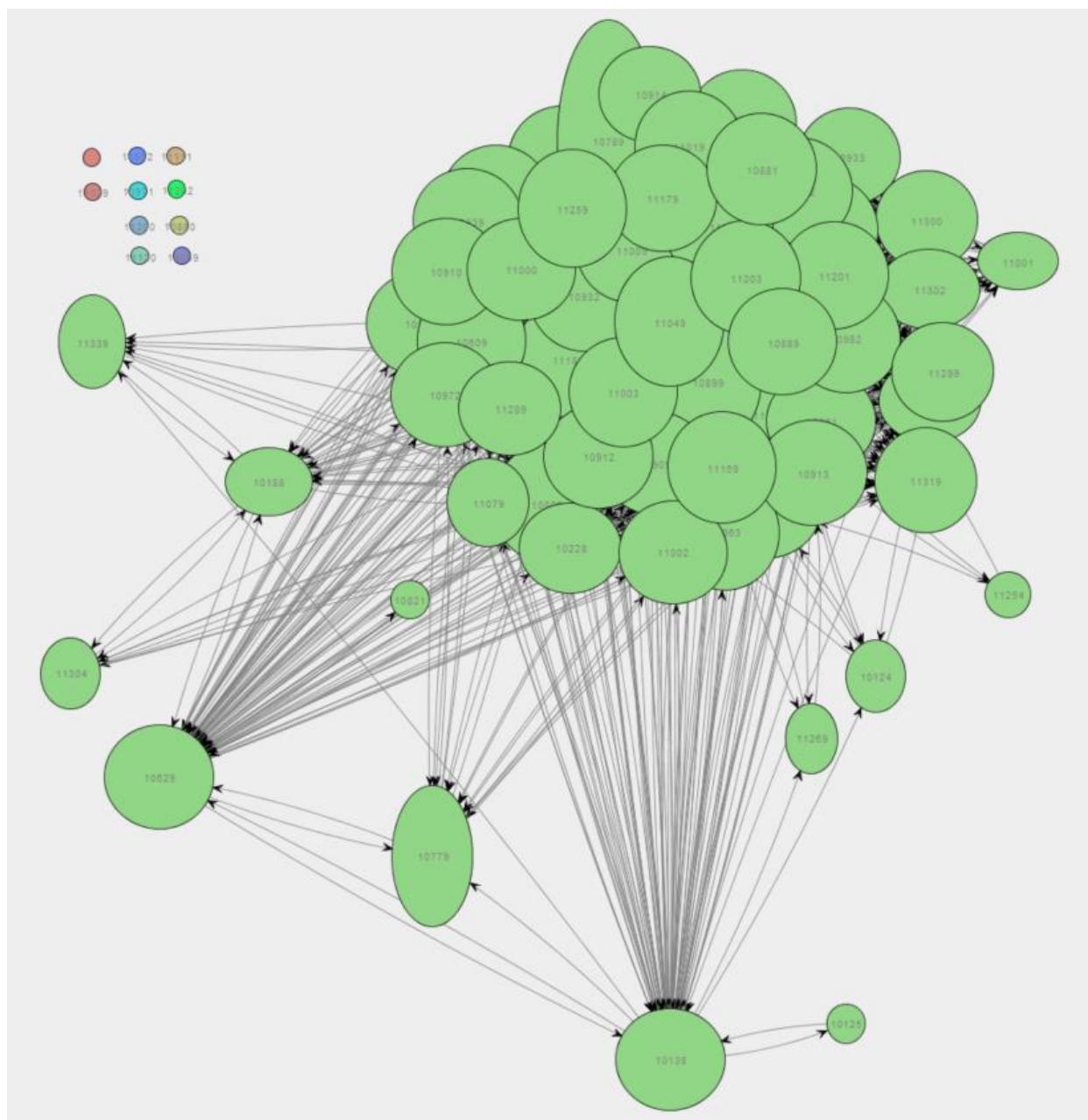


Figura 42: Handover of Work + sketch by degree ratio + size by ranking (ProM)

E inoltre isolando alcuni nodi attraverso la funzionalità “*picking*”, si possono mettere in evidenza alcuni aspetti come, ad esempio, la presenza di coppie di nodi che hanno un comportamento di mutua interazione (come ad esempio 10 629 con 10 138 e 10 779).

Tipologie:

- Nodi scollegati
- Comportamento globale (*closeness*)
- Comportamenti di mutua interazione (*betweenness*)

## 7.2 Mine for a Reassignment Social Network

Il secondo Plugin selezionato è il *Mine for a Reassignment Social Network*, la sua funzione principale è quella di evidenziare le deleghe di lavoro, questo tool risulta indispensabile per la comprensione di eventuali attività riassegnate ad altre risorse.

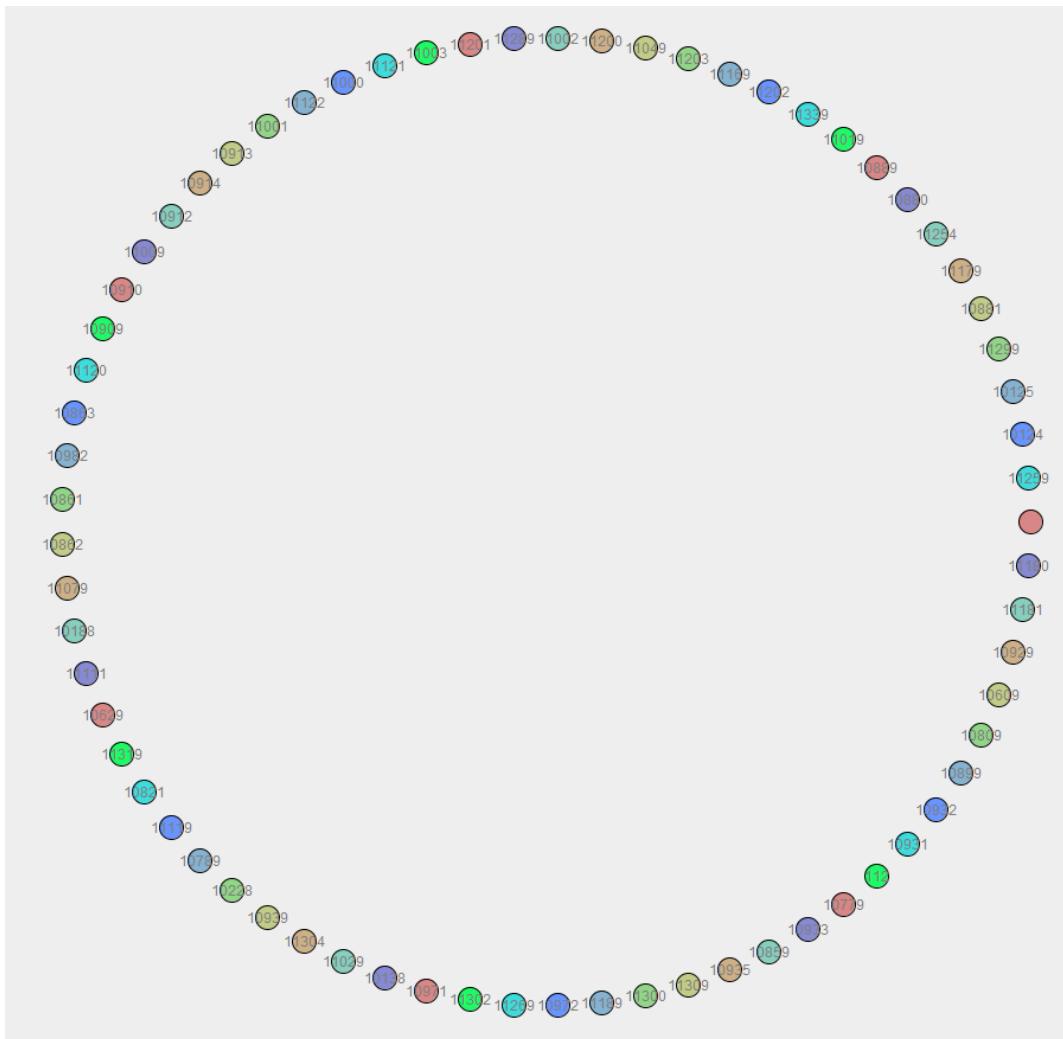


Figura 43: Mine for a Reassignment Social Network (ProM)

Essendo le risorse scollegate fra di loro, ciò sta a significare che non ci sono deleghe di lavoro tra le risorse, questo non è necessariamente un aspetto positivo in quanto il carico di lavoro non viene redistribuito correttamente.

### 7.3 Mine for a Similar-Task Social Network

In questo paragrafo è stato utilizzato il plugin “*Mine for a Similar-Task Social Network*”, il cui obiettivo è trovare cluster/gruppi di persone che lavorano agli stessi task.

Per misurare la “distanza” tra due risorse si è scelto di utilizzare il coefficiente di correlazione, in quanto permette una visualizzazione migliore per lo scopo per cui si è prefissati.

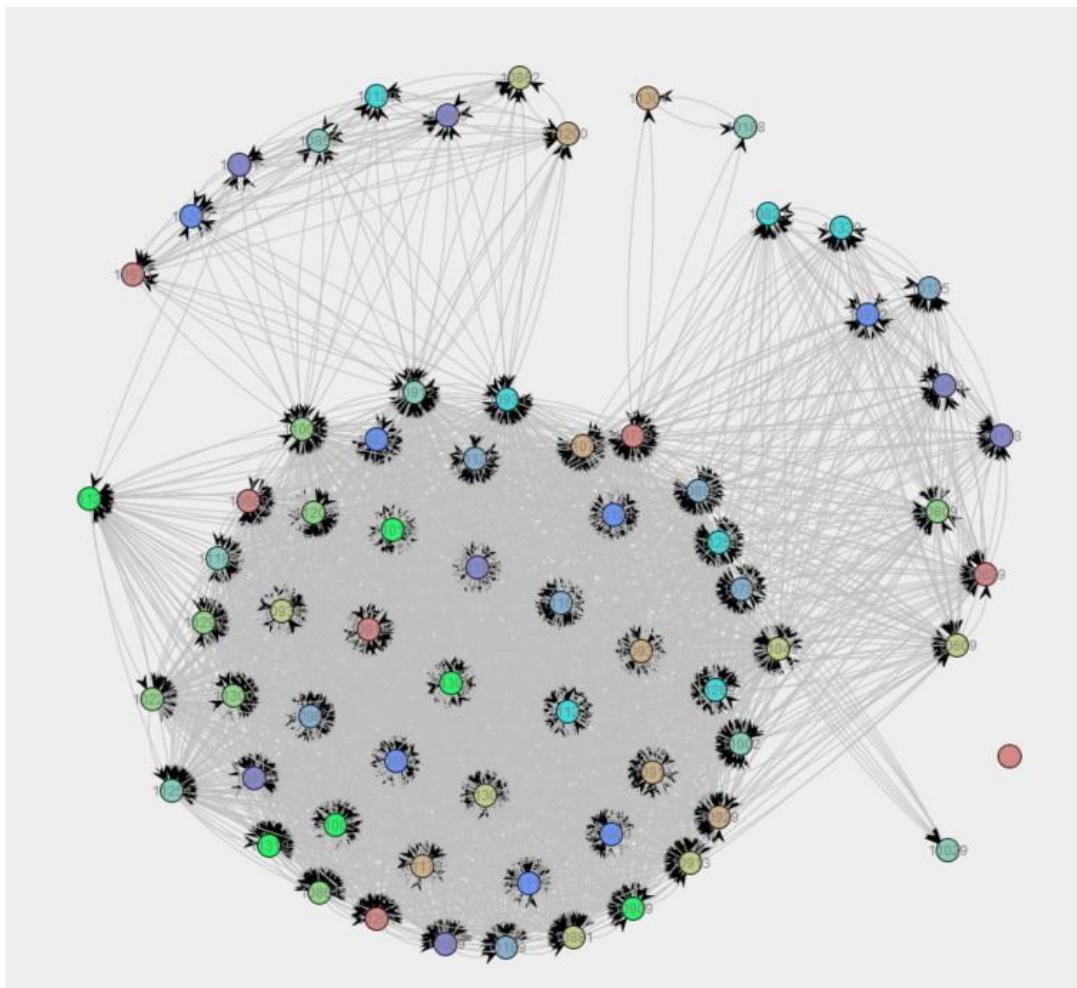


Figura 44: *Mine for a Similar-Task Social Network (ProM)*

Si può notare come solo una risorsa, che non è contrassegnata da un identificativo, non sia collegata agli altri nodi del grafo. È ipotizzabile che tale risorsa sia esterna all’istituto finanziario, o sia un supervisore di alcuni *task* di tipo “W”.

## 7.4 Mine for a Subcontracting Social Network

Grazie al *plug in* “Mine for a Subcontracting Social Network” è possibile identificare tutte quelle risorse che fanno un passaggio di Handover rimandandolo indietro. In gergo indica quante volte un individuo, ricevendo il lavoro da un soggetto, lo restituisce a quello stesso soggetto.

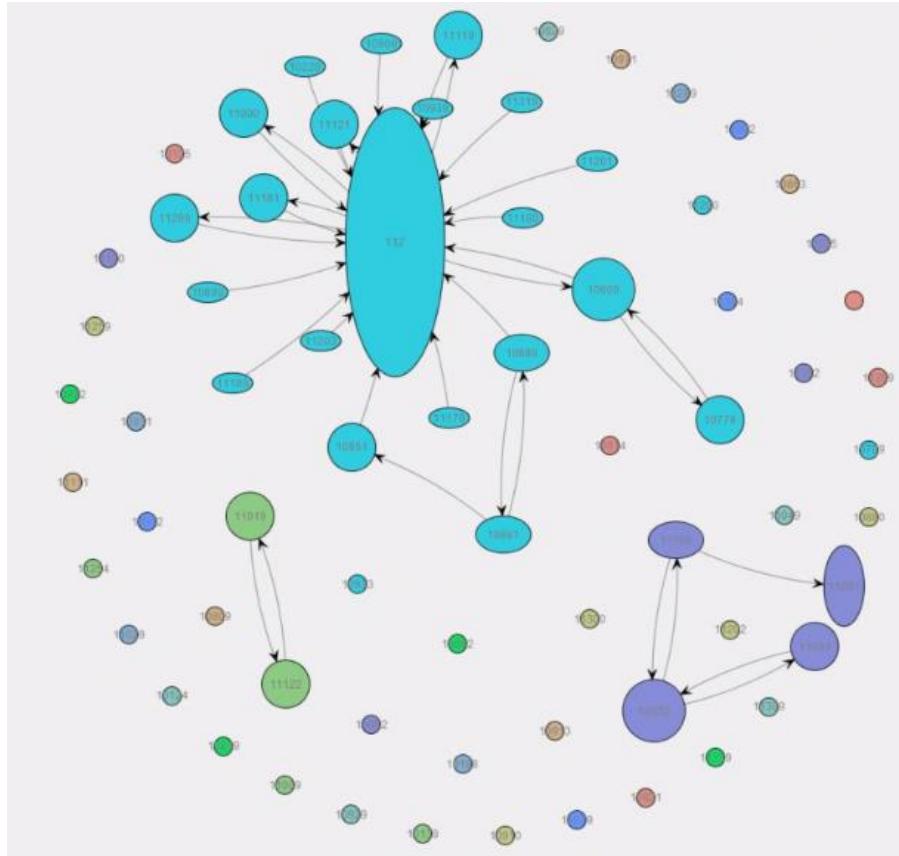


Figura 45: Mine for a Subcontracting Social Network (ProM)

Si identificano tre gruppi di diversi colori:

- *Celeste*: l’attività 112 è quella centrale e inoltre dalla forma si evidenzia che il numero di archi in ingresso è nettamente maggiore di quelli in uscita, inoltre alcune risorse non fanno passaggio di Handover;
- *Viola*: le risorse interessate sono quattro, di cui solo una non fa passaggio di Handover;
- *Verde*: le risorse sono due e sono perfettamente bilanciate fra di loro.

## 7.5 Mine for a Working-Together Social Network

Questo plug-in conta quanto frequentemente due o più risorse eseguono attività nello stesso case. Se le risorse lavorano insieme sugli stessi case, allora si avrà una relazione più forte tra le due risorse rispetto ad altre che lavorano insieme raramente. Si può notare come tutte le risorse abbiano una forte relazione fra di loro, eccezion fatta per la risorsa senza identificativo, ciò va ad avvalorare l'ipotesi precedentemente esposta.

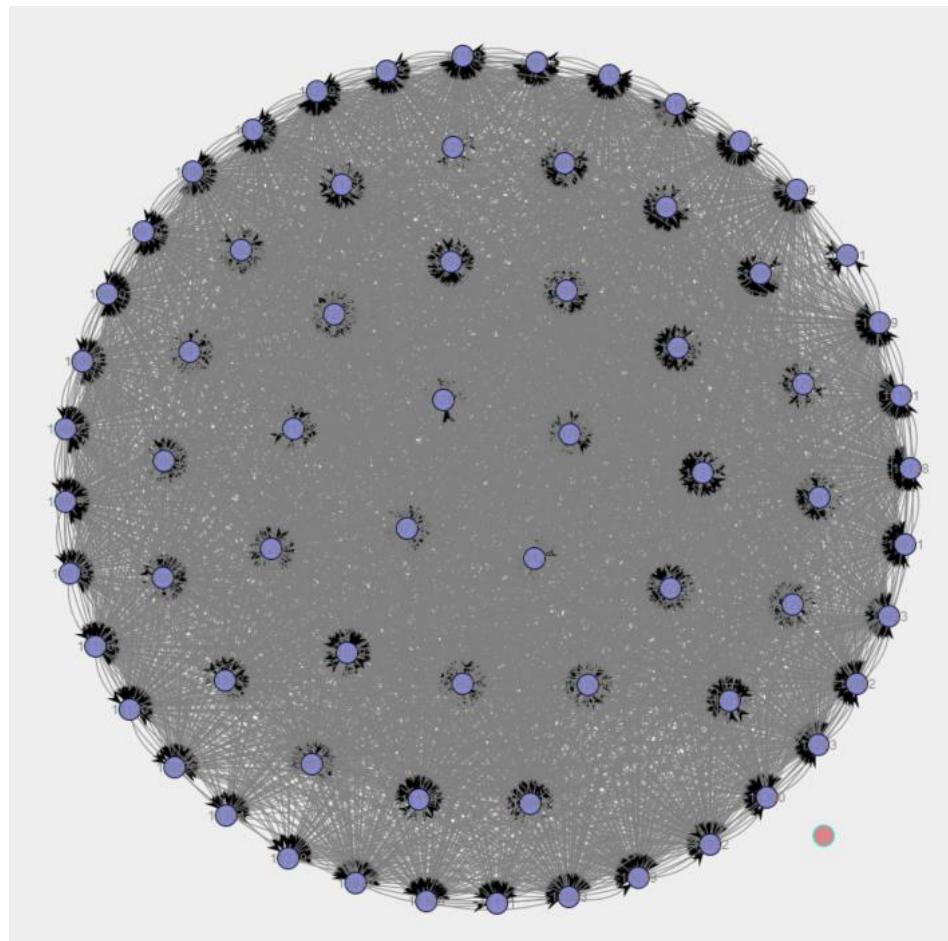


Figura 46: Mine for Working-Together Social Network

## 7.6 Dotted chart

In questo paragrafo ci si è focalizzati sul lavoro delle risorse svolto nel tempo, cercando di evidenziare tutte quelle risorse che sono state sovraccaricate di lavoro o al contrario sono state impiegate in un periodo di tempo saltuario.

In particolare, le seguenti risorse risultano poco impiegate nel tempo:

- 11001
- 11079
- 11269
- 11339
- 10124
- 10125
- 10914
- 10933
- 10935
- 10971

Mentre, altre lavorano saltuariamente (solo nei primi mesi):

- 11029
- 11111
- 11120
- 10859
- 10880

E solo negli ultimi mesi:

- 11289
- 11299
- 11300
- 11302
- 11304
- 11309
- 11319



Figura 47: Dotted Chart delle risorse nel tempo (ProM)

Di seguito, vengono invece riportate le analisi delle *Dotted Chart* riguardanti le attività che svolgono le singole risorse:

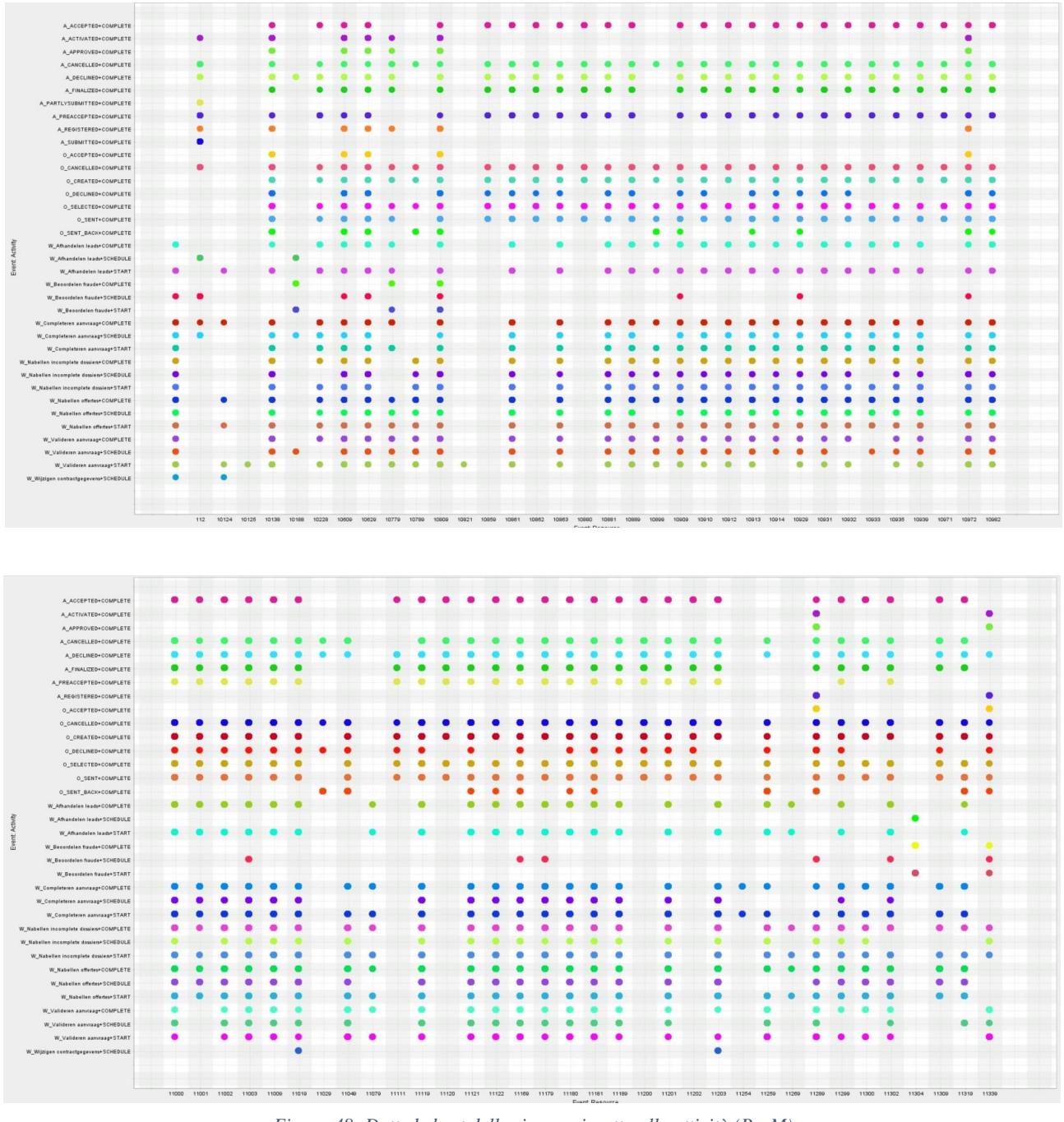


Figura 48: Dotted chart delle risorse rispetto alle attività (ProM)

Per poter effettuare l'analisi delle risorse con *Dotted Chart* su ProM, è stato necessario innanzitutto estrarre lo stesso dataset da Disco per bypassare un bug di ProM sulla visualizzazione delle risorse e in seguito è stato ineluttabile dividere la totalità delle risorse in due sottogruppi al fine di poterle catturare tutte e visualizzarne i nomi sugli assi.

## 7.7 Analisi risorsa 112 (Celonis)

È la risorsa che esegue più eventi (45 687) e con le seguenti medie:

- 274 eventi per giorno
- 7 attività per giorno
- 207 ore di completamento per lavoro

Le sue attività più frequenti sono:

- *A\_PARTLYSUBMITTED*
- *A\_SUBMITTED*
- *W\_Completeren aanvraag*
- *A\_PREACCEPTED*

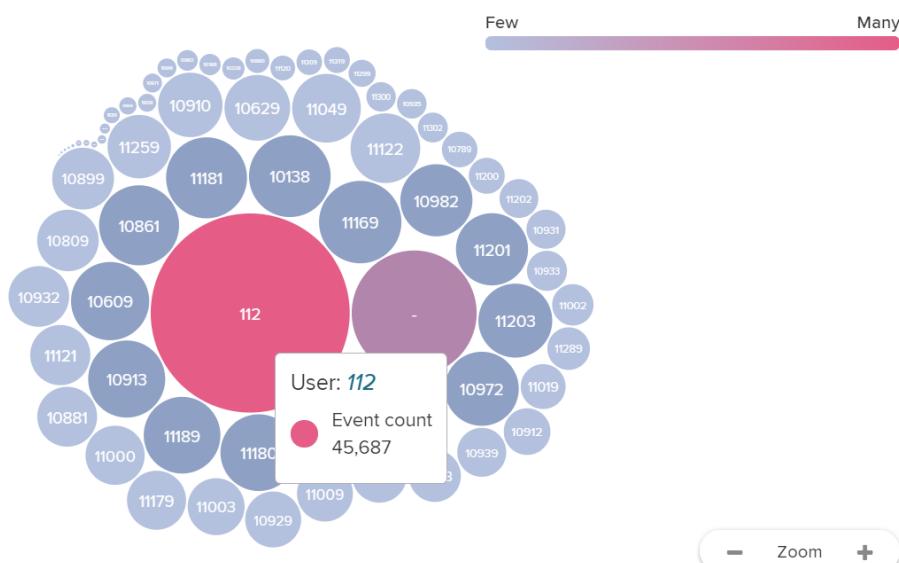


Figura 49: Numero degli eventi della risorsa 112 (Celonis)

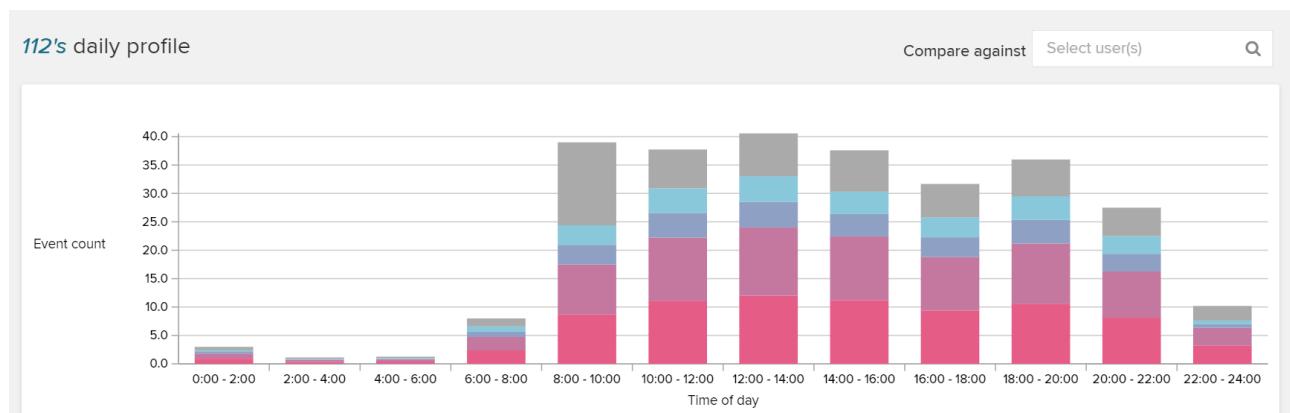


Figura 50: Distribuzione giornaliera degli eventi della risorsa 112 (Celonis)

Dall'orario lavorativo giornaliero si può stabilire che si tratta di una risorsa automatizzata, in quanto gestisce richieste in qualsiasi fascia oraria.

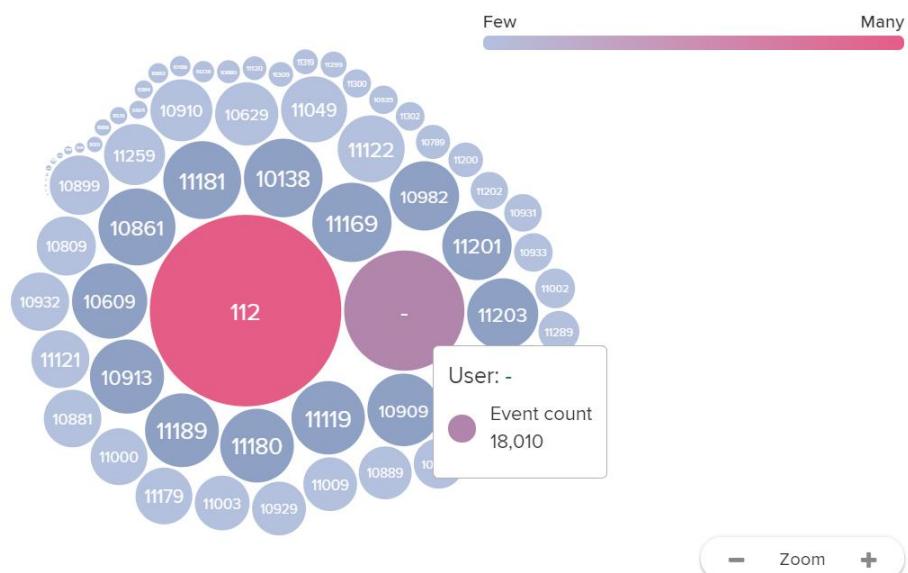
## 7.8 Analisi risorsa senza nome (Celonis)

È la risorsa che esegue più eventi (18 010) dopo la risorsa 112 e con le seguenti medie:

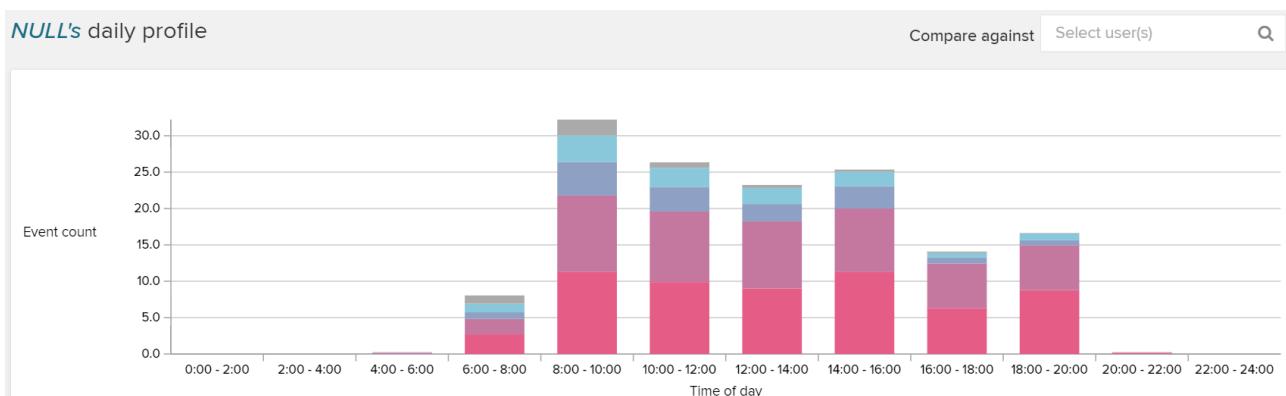
- 146 eventi per giorno
  - 3 attività per giorno
  - 406,9 ore di completamento per lavoro

Le sue attività più frequenti sono:

- *W\_Nabellen offertes*
  - *W\_Completeren aanvraag*
  - *W\_Nabellen incomplete dossiers*
  - *W\_Afhandelen leads*



*Figura 51: Numero degli eventi della risorsa NULL (Celonis)*



*Figura 52: Distribuzione giornaliera degli eventi della risorsa NULL (Celonis)*

Dall'orario lavorativo giornaliero si può stabilire che si tratta di una risorsa non automatizzata, in quanto gestisce richieste in fasce orarie che rappresentare realisticamente degli orari di lavoro.

## 8 Conclusioni

L’obiettivo del presente elaborato è stato quello di analizzare il log della BPI Challenge 2012 relativo alle richieste di prestito/scoperto presso un istituto finanziario olandese. L’analisi è stata svolta tramite diversi step, ognuno dei quali ha permesso di trarre diverse considerazioni. In primo luogo, vista la complessità del log, si è scelto di fare un’analisi comparativa tra i diversi *sublog* ottenuti e il log preso per intero; ciò però non ha sempre messo in luce criticità. Difatti, essendo i *sublog* dei processi secondari intrecciati, era ipotizzabile sin dall’inizio. Inoltre, dal processo preso in analisi, ci si è accorti che il numero delle varianti è estremamente elevato e molte di queste rappresentano un solo caso: le cause possono essere ricondotte come detto prima all’intrecciarsi di processi secondari, a richieste diverse da cliente a cliente che quindi generano tracce variegate, oppure a problematiche riscontrate in itinere (come, ad esempio, il controllo della frode). Inoltre, il comportamento principale del processo corrisponde ad un *path* di rifiuto delle richieste. Le richieste e il loro processamento sono comunque lineari, riscontrando solo alcuni casi particolari nella durata totale. In merito alla *Conformance* sono state riscontrate alcune deviazioni rilevanti quali ad esempio “*A\_PREACCEPTED*”, “*W\_Completeren\_aanvraag*”, “*A\_CANCELLED*” che, se incluse, portano ad un miglioramento del 17% in termini di conformità. Dall’analisi delle performance sono stati riscontrati inoltre alcuni colli di bottiglia, in particolare le sequenze di task “*W\_Nabellen\_offertes-COMPLETE* -> *W\_Nabellen\_offertes-COMPLETE*”, “*W\_Completeren\_aanvraag-COMPLETE* -> *W\_Nabellen\_offertes-COMPLETE*” e “*W\_Completeren\_aanvraag-COMPLETE* -> *W\_Complteren\_aanvraag-START*”. Analizzando le risorse invece, si nota che non tutte lavorano costantemente: alcune infatti lavorano solo in alcuni periodi specifici, altre lavorano per un periodo di tempo sotto la media (un caso isolato è la risorsa 10821); inoltre, due risorse in particolare (con id=112 e NULL) sono più frequenti rispetto alle altre. È ipotizzabile che la risorsa NULL sia esterna all’istituto finanziario, o sia un supervisore di alcuni *task* di tipo “*W*”.

Dal lavoro svolto, sono emerse alcune problematiche e alcune considerazioni in merito. In primo luogo, in relazione alla mappa generale del processo, si evidenziano alcune ridondanze nei task che, seppur non inficino nella durata complessiva, potrebbero essere eliminate (come, ad esempio, *A\_PARTLYSUBMITTED*). Si suggerisce inoltre di migliorare i controlli sui dati inseriti dal cliente prima che la richiesta venga inviata, in quanto in molti *case* sono presenti inconsistenze nei dati. Sono stati riscontrati oltretutto tempi di attesa lunghi per quanto riguarda l’interazione banca-cliente per la contrattazione dell’offerta; infatti, la banca è costretta a contattare il cliente più volte al fine di ricevere una risposta positiva alla sua offerta. Le richieste prese in carico sono perlopiù rifiutate e infatti, i task *A\_DECLINED* e *A\_CANCELLED* sono presenti nella maggior parte dei case; invece, gli stati finali di successo (“*A\_REGISTERED*”, “*A\_APPROVED*”, “*A\_ACTIVATED*”) per l’accettazione della richiesta sono presenti nel 17% dei case (solo 2246 su 13087 case). Per quanto riguarda le risorse invece è consigliabile ridurne il numero, specializzandone alcune, al fine di ridurre i costi del personale ed evitando che ognuno svolga più operazioni contemporaneamente. Le risorse più frequenti invece, denominate con id=112 e NULL, dovrebbero essere potenziate in quanto risorse chiavi dell’intero processo.