Noname manuscript No.
(will be inserted by the editor)

Applying process mining techniques in a real healthcare case study

M. Mecella · F. Covino · A. Marrella · S. Agostinelli

Received: date / Accepted: date

Abstract The healthcare organizations are under increasing pressure to improve productivity, gain competitive advantage and reduce costs. For this reason, healthcare organizations, such as hospitals try to streamline their processes. In this paper we demonstrate the applicability of process mining in the healthcare domain, using a real case study of '*San Carlo di Nancy*' hospital in Rome (GVM Group). We apply process mining techniques to obtain meaningful knowledge about the patient careflows from so-called event logs, obtained from raw data of hospital information systems. We analyzed these logs using the ProM framework from three different perspectives: the control flow perspective, the organizational perspective and the timing perspective. The results show that process mining can be used to provide new insights that facilitate the improvement of existing careflows.

Keywords Process mining · Healthcare · ProM

M. Mecella
Professor in Engineering in Computer Science at Sapienza University of Rome
E-mail: mecella@dis.uniroma1.it

A. Marrella
Postdoctoral research fellow in Computer Science and Engineering at Sapienza University of Rome
E-mail: marrella@dis.uniroma1.it

F. Covino
CMMI Business Analyst at ENAV
E-mail: federico.covino@gmail.com

S. Agostinelli
Graduated in Computer Engineering at Sapienza University of Rome
E-mail: simone.agostinelli.sa@gmail.com

1 Introduction

Nowdays, the hospitals try to streamline their processes in order to deliver high quality care while at the same time improving revenues and reducing costs. More and more pressure is put on hospitals to work in the most efficient way as possible, whereas in the future, an increase in the demand for care is expected. A complicating factor is that healthcare is characterized by highly complex and extremely flexible patient care processes, also referred to as '*careflows*'. In healthcare organisations, a wide range of processes with different characteristics and requirements are daily managed and executed. The delivery of complex care may involve several departments and organisations, and requires an active collaboration between different professionals and practitioners having heterogeneus skills. Healthcare is thus widely recognised as one of the most promising, yet challenging, domains for the adoption of process-oriented solutions. We demonstrates the applicability of process mining in the healthcare domain, using a real case study of '*San Carlo di Nancy*' hospital in Rome (GVM Group). We apply process mining techniques to obtain meaningful knowledge about the patient careflows from so-called '*event logs*' obtained from raw data of hospital information systems. Process mining aims at extracting process knowledge from that logs in order to discover, for example, both typical paths followed by particular groups of patients and strong collaboration between different hospitalization wards. We analyzed the different careflows both under the control flow perspective (emphasizing the differences between the process models obtained from different cluster of patients), the organizational perspective (looking at the social networks we were able to discover the relationship between the resources of the patient careflows) and the performance perspective (looking at the timing perspective of different activities performed by the patients we were able to discover bottlenecks in the patient careflows). In order to do so, we extracted the event logs from the raw datasets of '*San Carlo di Nancy*' hospital and we analyzed them using *ProM*: the process mining framework. The datasets in question are the following:

- *Ambulatori* (outpatient clinic): each row stores the information about a single healthcare service.
- *Pronto soccorso* (emergency room): each row represents a single emergency room activity.
- *Ricoveri* (hospitalizations): each row represents a single hospitalization taken by a patient.

These three datasets contain raw data about patients treated in both year 2016 and May 2017 for which all the treatment activities have been recorded. We did not use any a-priori knowledge about the careflows of the patients of '*San Carlo di Nancy*' hospital and did not have any process model at hand. The data analyzed are the standard ones of the National Health Service (Servizio Sanitario Nazionale) that the hospitals interchanged with the Regional Authorities (Enti Regione). Therefore the presented analysis can be replicated nationwide.

2 Analysis

In this section, we discuss the results obtained by applying process mining techniques (process discovery, social network analysis and performance analysis) on the datasets of '*San Carlo di Nancy*' hospital already '*converted*' to event logs: '*Ambulatori*', '*Pronto soccorso*' and '*Ricoveri*'. The process mining techniques we used for discovering the process model of the totality of patients (blueprint) are: the *Alpha Miner* algorithm, the *Heuristic Miner* algorithm and the *Inductive Miner* algorithm. In order to establish which process model out of three is representing better the behaviour observed in the event log, and therefore which mining algorithm is the best, we computed the four quality metrics replay fitness, precision, generalization and simplicity. ProM 6.7 allows to compute replay fitness, precision and generalization only. Therefore simplicity is calculated on the basis of this metric: $\#activities + \#splits + \#joins$. The grater is the value, the less is the simplicity of the process model. At the end, the best algorithm will be chosen on the basis of the just mentioned quality metrics. Since the inductive visual miner is high parametrisable (i.e., we can set a-priori the replay fitness we would like to reach in order to obtain a process model in function of this fitness value) therefore it's the algorithm with the best compromise between the just mentioned quality metrics. For this reason, we picked the inductive visual miner as process discovery technique. After discovered the blueprint of all the three datasets, we have performed ad-hoc analysis for these set of clusters:

- foreign patients [1] vs. italian patients [2]
- patients with an age lower or equal than 25 years (resident in Latium)[3] vs. patients with an age greater than 25 years (resident in Latium) [4]

Only for '*Ricoveri*' dataset we have performed a further in depth analysis for these specialist branches: *chirurgia generale*, *medicina generale* and *ortopedia and traumatologia* looking at both pre-operative hospitalization time (the time a patient is waiting for the surgical intervention) and post-operative hospitalization time (the time a patient is waiting for being discharged).

2.1 Ambulatori dataset

The blueprint in terms of process model is the following:

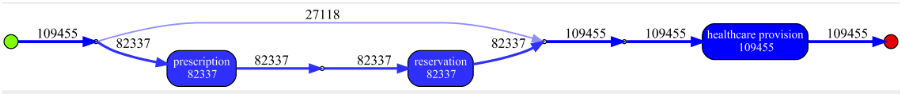


Fig. 1 Clinical path (Totality of patients)

Looking at the figure, we can see that the most of the patients takes the medical prescription and asks for a reservation before the healthcare provision. From the organizational point of view, the social network derived is as follows:

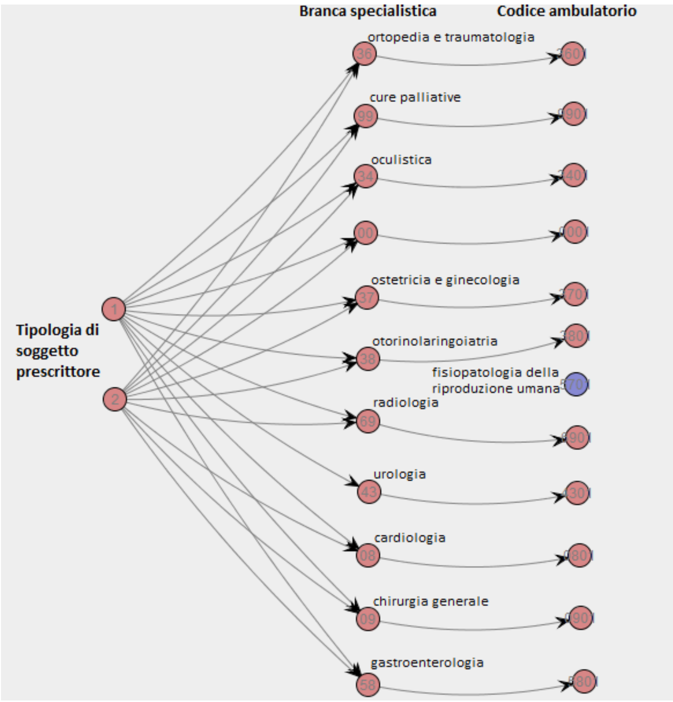


Fig. 2 Social network (Totality of patients)

From the picture points out that the types of prescriptive subjects involved in prescriptions are both '*1*' (medico di medicina generale) and '*2*' (medico specialista dipendente da struttura pubblica). They are connected with the specialist branches accredited for the healthcare provision, since the prescriptive subject chooses the specialist branch when he compiles the medical prescription. Notice that the outpatient clinic '*5701*' (Fisiopatologia della riproduzione umana) is not involved in any social relation, because all the patients go in that outpatient clinic without prescription. Indeed the process model followed by these patients is composed by only one activity, the ‘*healthcare provision*’ activity:

Fig. 3 Fisiopatologia della riproduzione umana

On the other hand, regarding the performance perspective we have obtained:

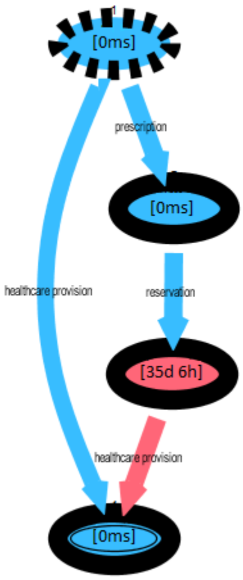


Fig. 4 Transition System (Totality of patients)

We can see that the most of time, from the sojourn time point of view, is spent on the activity ‘*reservation*’ where the patient is waiting for the healthcare provision. The standard deviation is very high, meaning that the patients are spread out over a wide range of time, therefore the mean is not significant at all. Then, we have repeated the same analysis for different cluster of patients. We started with both the clusters [1] and [2] obtaining the following process models:

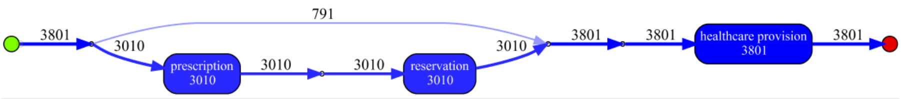


Fig. 5 Clinical path (Foreign patients)

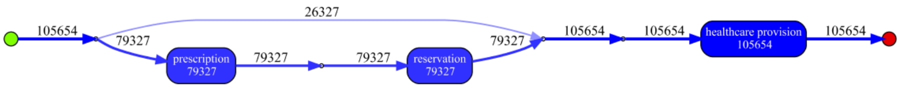


Fig. 6 Clinical path (Italian patients)

Both the process models are identical to the blueprint, the only difference is on the number of process executions (3801 for [1] and 105654 for [2]). Furthermore, looking at both the figures, we can see that the most of both foreign and italian patients take the medical prescription and ask for a reservation before the healthcare provision. From the organizational point of view, the social networks derived from both the clusters [1] and [2] are as follows:

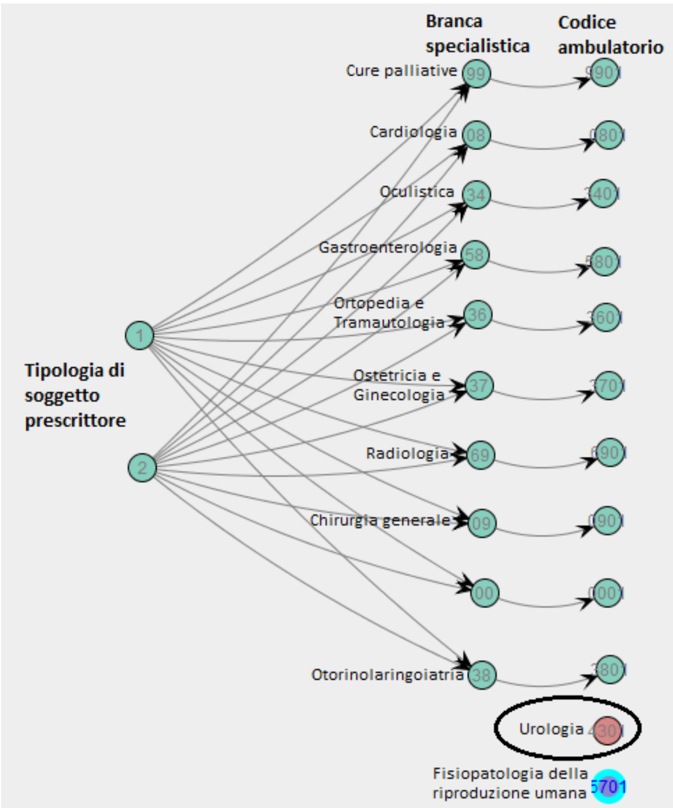


Fig. 7 Social Network (Foreign patients)

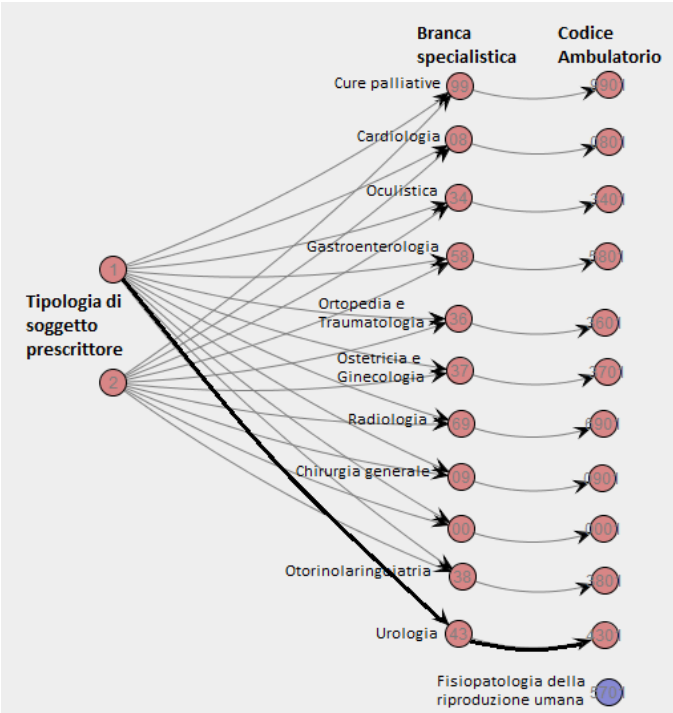


Fig. 8 Social Network (Italian patients)

Looking at the pictures, the social networks of the patients belonging from both the clusters [1] and [2] involve the same resources as the blueprint social network. The only difference is on the outpatient clinic '*4301*' (Urologia): the foreign patients go in this outpatient clinic without medical prescription instead the italian patients with a medical prescription compiled only from '*1*' (medico di medicina generale) .This means all the foreign patients going in '*4301*' perform only one activity, the '*health-care provision*' activity:

Fig. 9 Urologia (Foreign patients)

Looking at the performance perspective, in particular at the *sojourn time* point of view, we obtained:

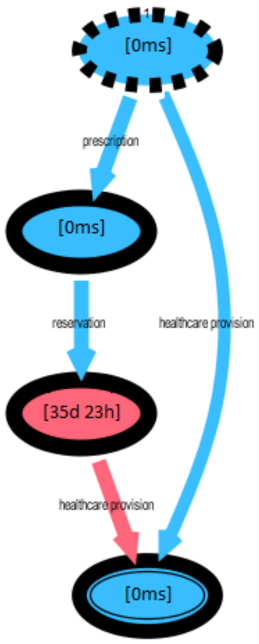


Fig. 10 Transition system (Foreign patients)

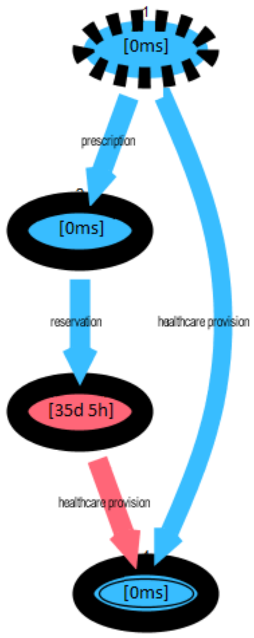


Fig. 11 Transition system (Italian patients)

At first glance, looking at the two transition systems we can see that the states associated with activities belonging to the cluster [2] have a border ticker than the states associated with activities belonging to the cluster [1], since the cluster of foreign patients contains a number of patient's cases strictly lower than the cluster of italian patients. The same argument applies to arcs. Moreover, we can see that the most of time, for both the clusters [1] and [2], is spent on the activity '*reservation*' where a patient is waiting for the healthcare provision. The avg sojourn time of both the clusters is equal (more or less) to the one of the blueprint.
Now we can consider another set of clusters: patients belonging both to [3] and [4]. The derived process models are the following:

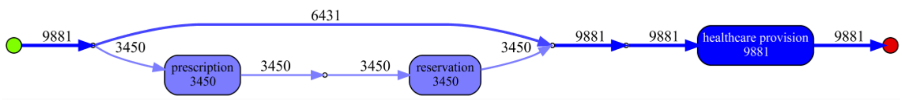


Fig. 12 Clinical path (≤25 years)

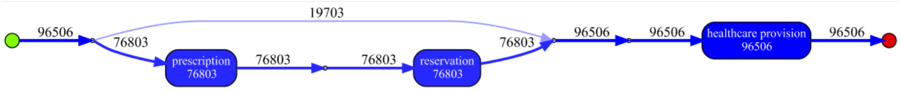


Fig. 13 Clinical path (>25 years)

Both the process models are identical to the blueprint, but there are an interesting difference regarding the behaviour of the patients belonging to both the clusters [3] and [4], in particular regarding the most frequent path. The most frequent path taken by the patients with an age lower than 25 years (resident in Latium) is the one made up by the '*healthcare provision*' activity (they go directly in outpatient clinic), instead, the most frequent path taken by the patients with an age grater or equal than 25 years (resident in Latium) is made up by the following activities: '*prescription*' $\rightarrow$ '*reservation*' $\rightarrow$ '*healthcare provision*' (they take the medical prescription, then they ask for a reservation for perfoming the healthcare provision). Another difference which points out is on the number of process executions (9881 for [3] and 96506 for [4]).

From the organizational point of view, the social networks derived from both the clusters [3] and [4] are as follows:

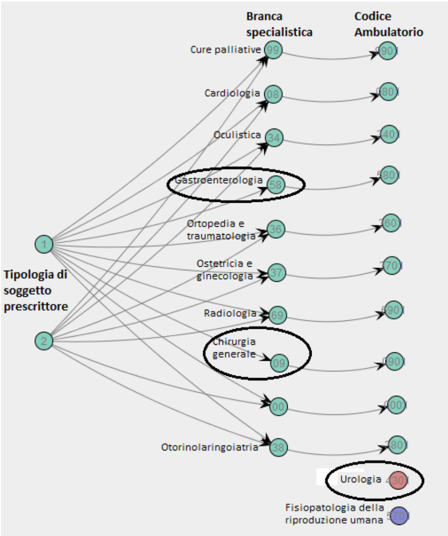


Fig. 14 Social network (≤25 years)

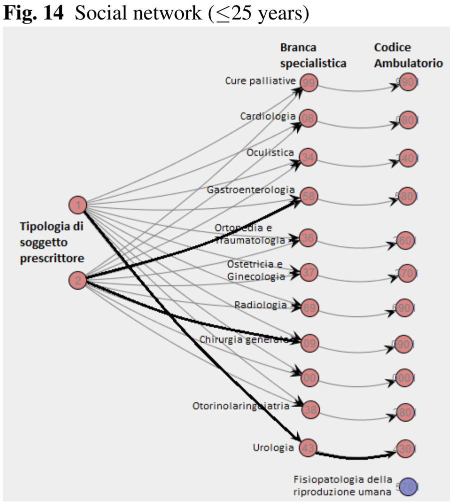


Fig. 15 Social network (>25 years)

Looking at the pictures, the social networks of the patients belonging from both the clusters [3] and [4] involve the same resources as the blueprint social network. The main difference is on the outpatient clinic '*4301*' (Urologia): the patients with an age lower than 25 years (resident in Latium) go in this outpatient clinic without medical prescription, instead, the patients with an age greater or equal than 25 years (resident in Latium) go in this outpatient clinic with a medical prescription compiled only by '*1*' (medico di medicina generale). This means all the patients belonging to the cluster [3] going in '*4301*' (Urologia) perform only one activity, the '*healthcare provision*' activity:

Fig. 16 Urologia (≤25 years)

Moreover, patients belonging from [3] go in both outpatient clinics '*5801*' (Gastroenterologia) and '*0901*' (Chirurgia generale) with medical prescriptions compiled only by '*1*' (medico di medicina generale). For patients belonging from [4] there are also cases of medical prescriptions compiled by '*2*' (medico specialista dipendente da struttura pubblica).
Looking at the performance perspective, in particular at the *sojourn time* point of view, we obtained:

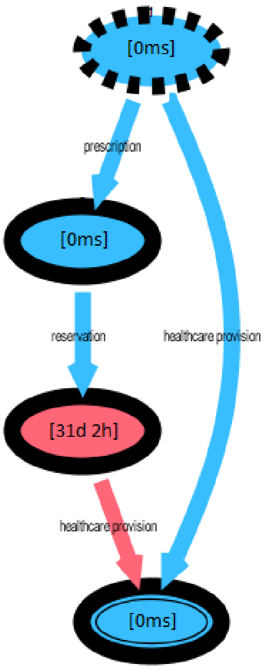


Fig. 17 Transition system (≤25 years)

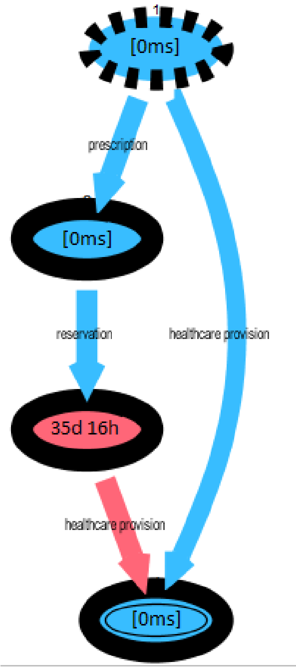


Fig. 18 Transition system (>25 years)

At first glance, looking at the two transition systems we can see that the states associated with activities belonging to the cluster [4] have a border ticker than the states associated with activities belonging to the cluster [3], since the cluster [3] contains a number of patient's cases strictly lower than the cluster [4]. The same argument applies to arcs. Moreover, we can see that the most of time, for both the clusters [3] and [4], is spent on the activity '*reservation*' where a patient is waiting for the healthcare provision. The avg sojourn time of both the clusters is equal (more or less) to the one of the blueprint.

3 Pronto soccorso dataset

The blueprint in terms of process model is the following:

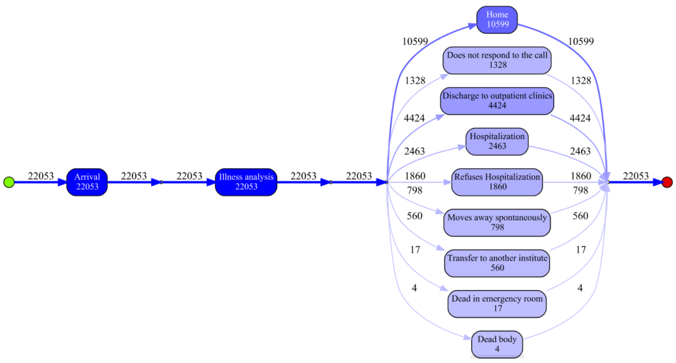


Fig. 19 Clinical path (Totality of patients)

Looking at the figure, we can see that a patient arrives in emergency room, then performs the problem diagnosis. From here, the patient can have several outcomes. It figures out that the majority of patients are sent home after the illness analysis. From the organizational point of view, the social network derived is as follows:

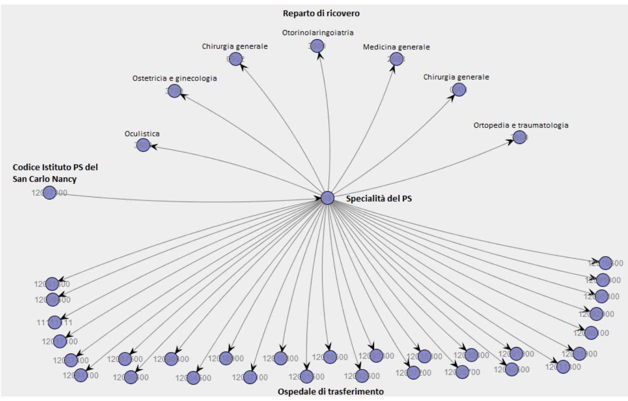


Fig. 20 Social network (Totality of patients)

From the social network points out a patient go in emergency room, that is, the one coded with '*12007300*' (San Carlo Nancy emergency room); then he is assigned to emergency room specialty '*99*' and from here he can be either transferred to another Institute or to another hospitalization ward. On the other hand, the transition system, under the *sojourn time* point of view is the following:

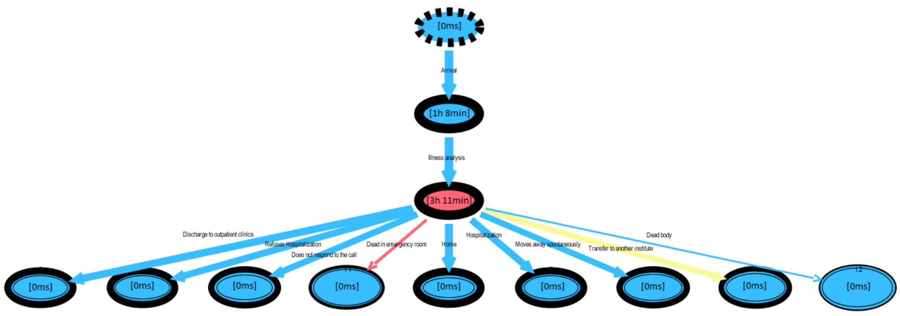


Fig. 21 Transition System (Totality of patients)

We can see that the most of time is spent on the activity '*illness analysis*'. Since, the standard deviation diverges from the different average times, meaning that the patients are spread out over a wide range of time,then the mean is not significant at all. For this reason, we have repeated the same analysis for different cluster of patients. We started with both the clusters of foreign [1] and the italian [2] patients, obtaining the following process models:

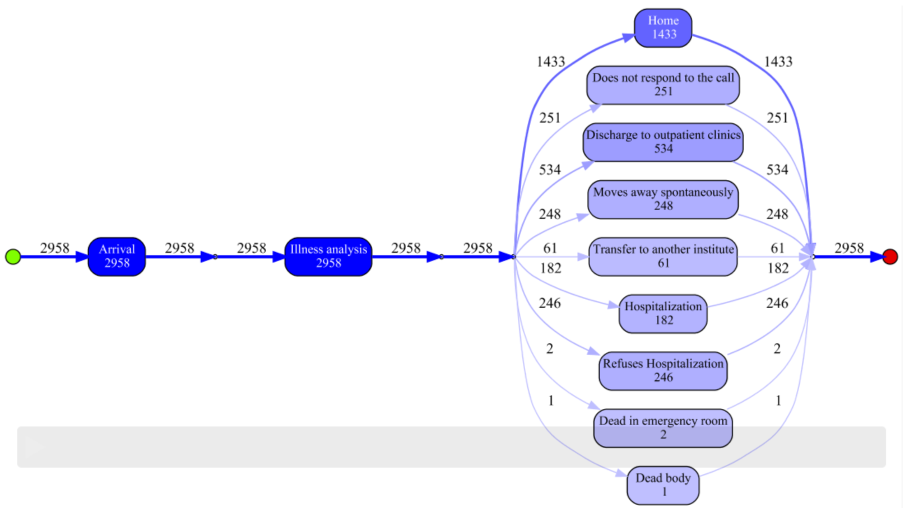


Fig. 22 Clinical path (Foreign patients)

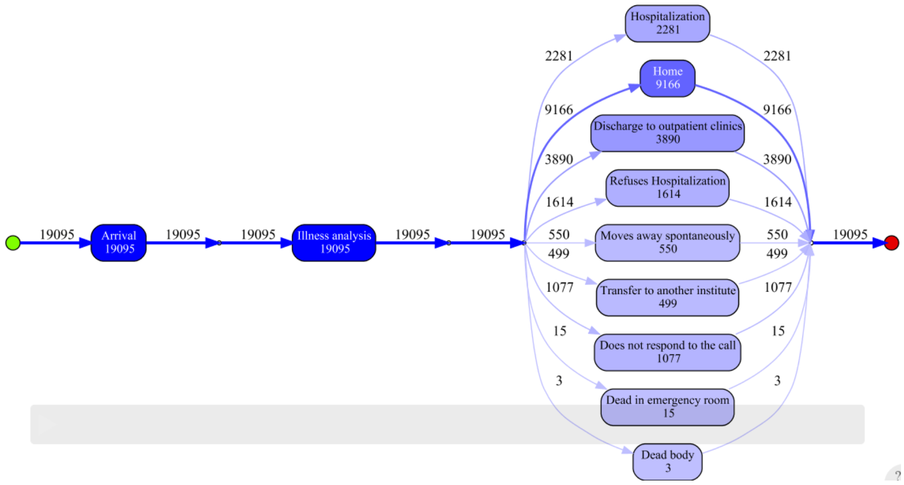


Fig. 23 Clinical path (Italian patients)

Both the process models are identical to the blueprint, the only difference is on the number of process executions (2958 for [1] and 19095 for [2]). Furthermore, looking at both the figures, we can see that the most of both foreign and italian patients after performed the illness analysis are sent home. From the organizational point of view, the social networks derived from both the clusters [1] and [2] are as follows:

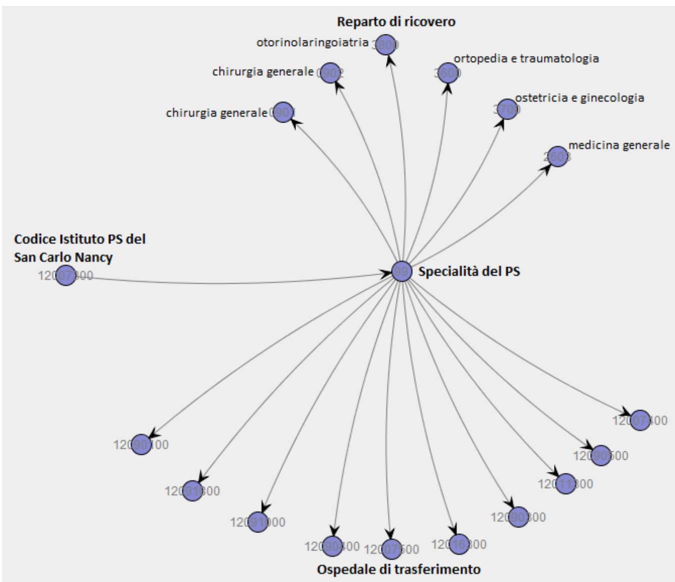


Fig. 24 Social network (Foreign patients)

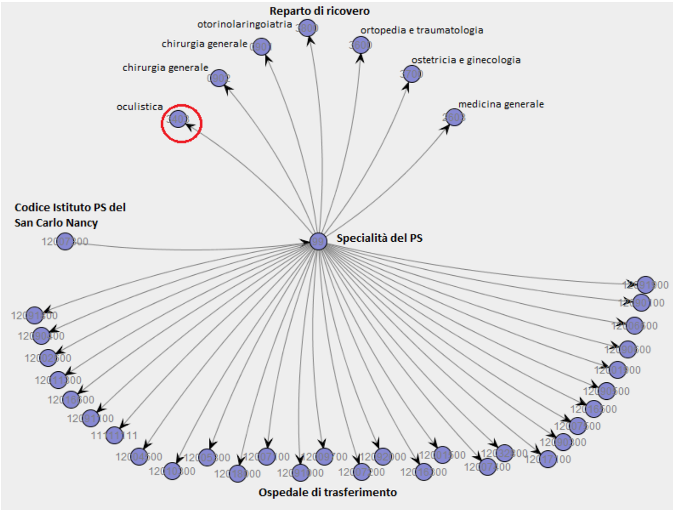


Fig. 25 Social network (Italian patients)

Looking at the pictures, the social networks of the patients belonging from both the clusters [1] and [2] are quite different. The main difference, which its pretty obvious, is that the number of resources involved in the cluster of the italian patients is more grater than the number of resources involved in the cluster of the foreign patients. In particular, italian patients are transferred in a greater number of Institutes. Moreover, there is the hospitalization ward '*3403*' (oculistica) belonging only to the cluster of italian patients. The transition systems, on the *sojourn time* point of view, are:

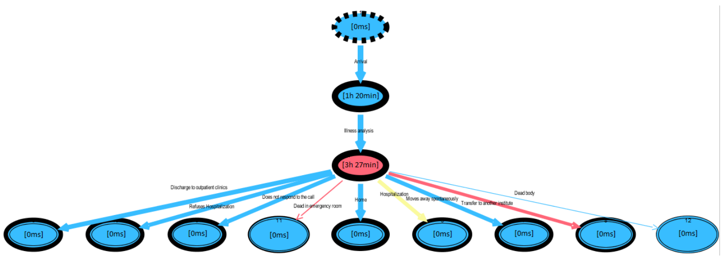


Fig. 26 Transition system (Foreign patients)

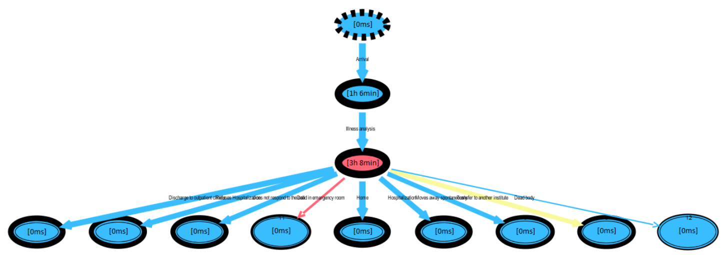


Fig. 27 Transition system (Italian patients)

At first glance, looking at the two transition systems we can see that the states associated with activities belonging to the cluster [2] have a border ticker than the states associated with activities belonging to the cluster [1], since the cluster [1] contains a number of patient's cases strictly lower than the cluster [2]. The same argument applies to arcs. Moreover, we can see that the most of time, for both the clusters [1] and [2] is spent on the activity '*illness analysis*' where the patient is waiting for the result of the diagnosis. The avg sojourn time of both the clusters is equal (more or less) to the one of the blueprint. Then, we have considered both the clusters of patients with an age lower or equal than 25 years (resident in Latium) [3] and the cluster of patients with an age greater than 25 years (resident in Latium) [4]. The process models obtained are the following:

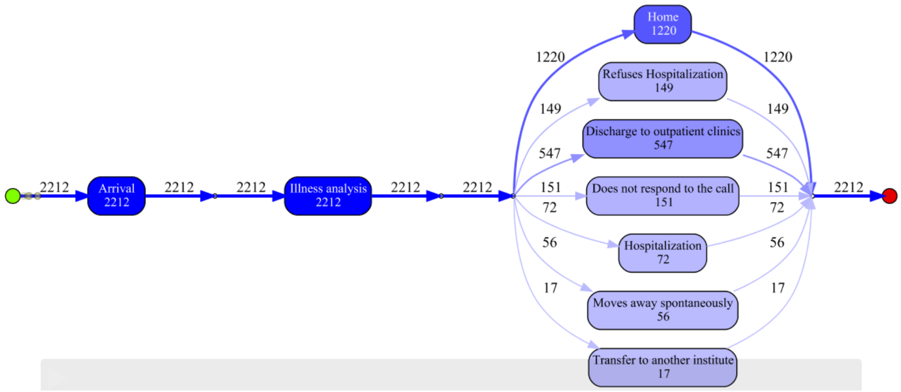


Fig. 28 Clinical path (≤25 years)

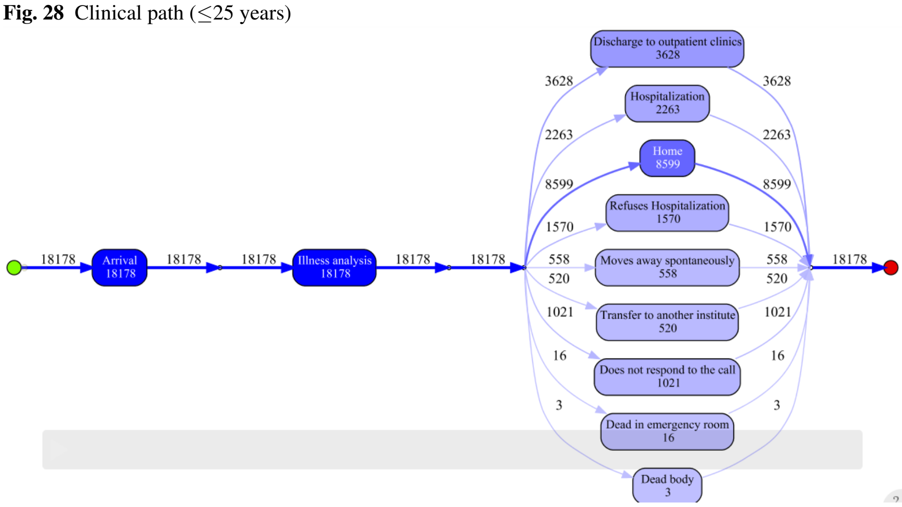


Fig. 29 Clinical path (>25 years)

Looking at the process model we can see that the patients with an age lower or equal than 25 years (resident in Latium) [3] does not have the activities: '*dead body*' and '*dead in emergency room*', with respect to the cluster of patients with an age greater than 25 years (resident in Latium) [4]. However, the behaviour of the patients belonging to these two clusters is pretty equal; also the most frequent path is the same: '*arrival*' ⇒ '*illness analysis*' ⇒ '*home*'. The derived social networks for both [3] and [4] are:

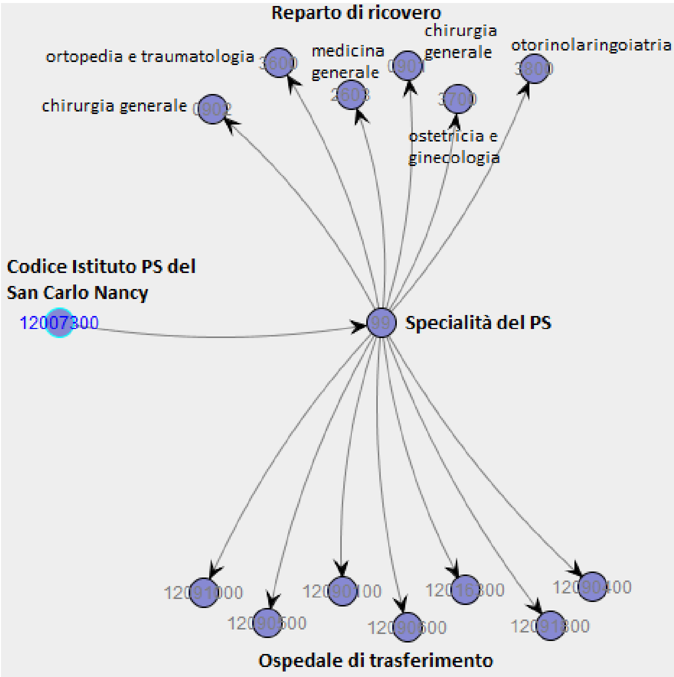


Fig. 30 Social network (≤25 years)

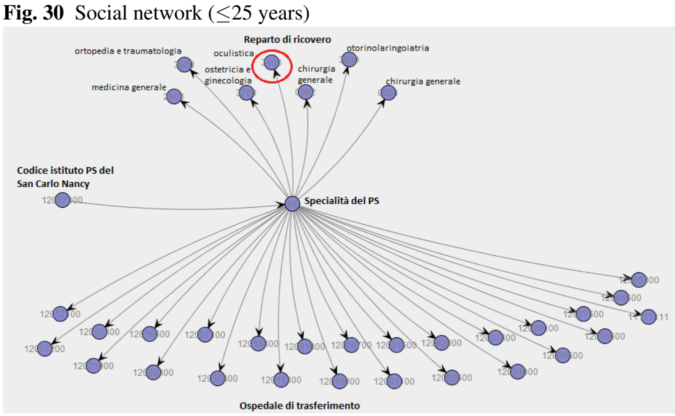


Fig. 31 Social network (>25 years)

The main difference, which its pretty obvious, is that the number of resources involved in the cluster of patients with an age greater than 25 years (resident in Latium)

[4] is more greater than the number of resources involved in the cluster of patients with an age lower or equal than 25 years (resident in Latium) [3]. In particular, the patients of the cluster [4] are transferred in a greater number of Institutes. Furthermore, there is the hospitalization ward '*3403*' (oculistica) belonging only to the cluster [4]. The transition systems, on the *sojourn time* point of view, are:

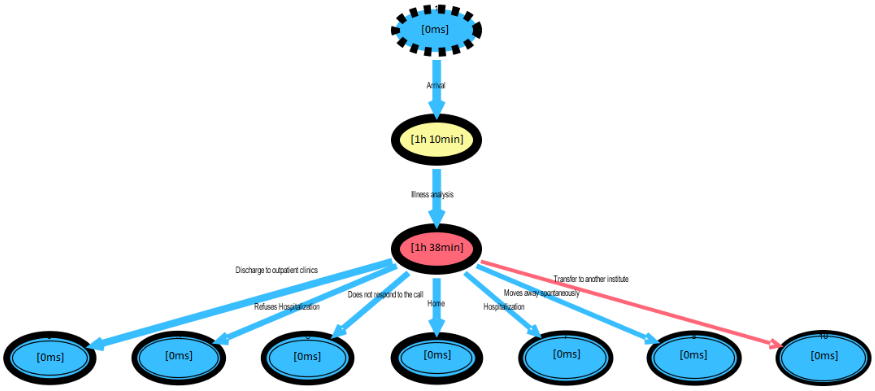


Fig. 32 Transition system (≤25 years)

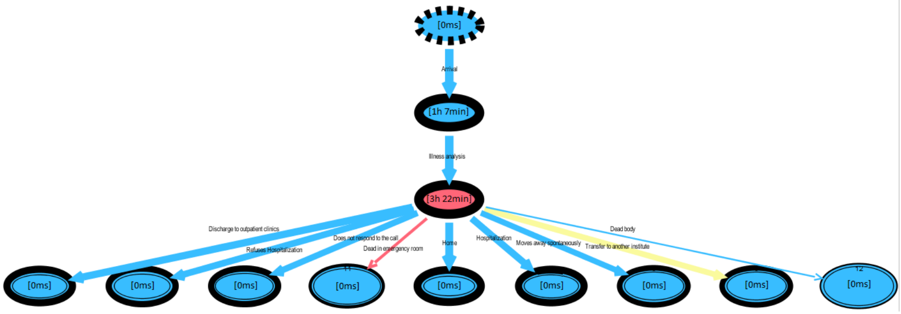


Fig. 33 Transition system (>25 years)

At first glance, looking at the two transition systems we can see that both states and arcs associated with activities belonging to the cluster [4] have a border ticker than the states associated with activities belonging to the cluster [3] (the reason is always the same). Moreover, we can see that the most of time, for both the clusters [3] and [4] is spent on the activity '*illness analysis*' where the patient is waiting for the result of the diagnosis.

4 Ricoveri dataset

The blueprint in terms of process model is the following:

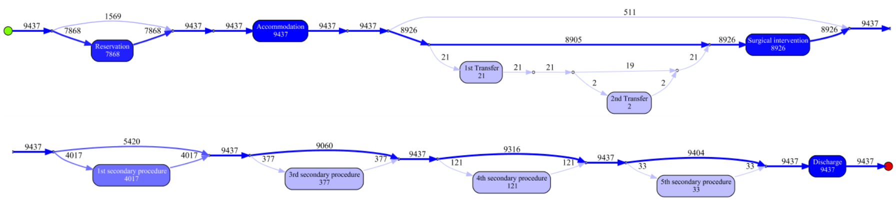


Fig. 34 Clinical path (Totality of Patients)

Looking at the figure, we can see that a patient can ask for a reservation, then he/she is accommodated in order to perform the surgical intervention. Notice that, before the surgical intervention, a patient can be transferred from a hospitalization ward to another one. However, the transfers could happen also after the surgical intervention and between the secondary procedure. This behavior is not captured by the process model, and therefore it contains deviations with respect to event log. After the surgical intervention a patient can be either directly discharged or submitted through a chain of secondary procedures that will end with the discharge of the patient. Notice that a patient may skip the surgical intervention.. From the organizational point of view, the derived social network is the following:

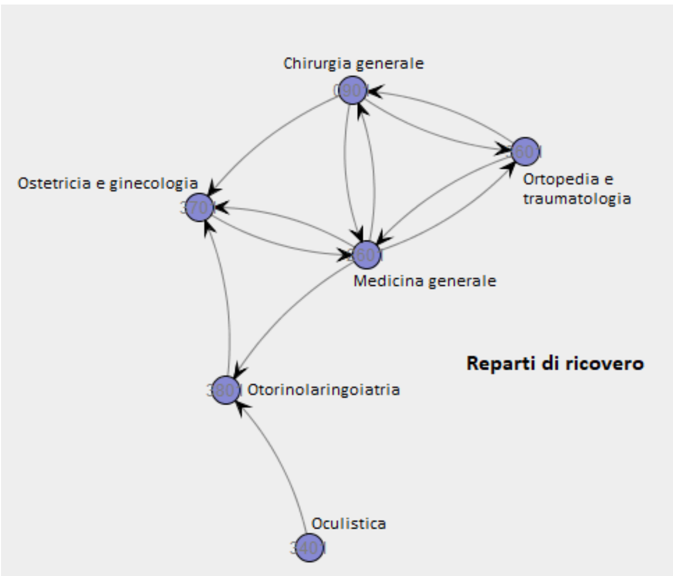


Fig. 35 Social Network (Totality of Patients)

The social network shows how the different hospitalization wards interact between each other. The semantic of the social network is the following: an arc from a hospitalization ward '*x*' to another hospitalization ward '*y*' means that exist at least a patient transferred from '*x*' to '*y*'. For example, if we take in consideration the hospitalization ward '*3401*' (oculistica), there is at least a patient that is transferred to the hospitalization wards '*3801*' (otorinolaringoiatria).
Regarding the performance perspective we have obtained:

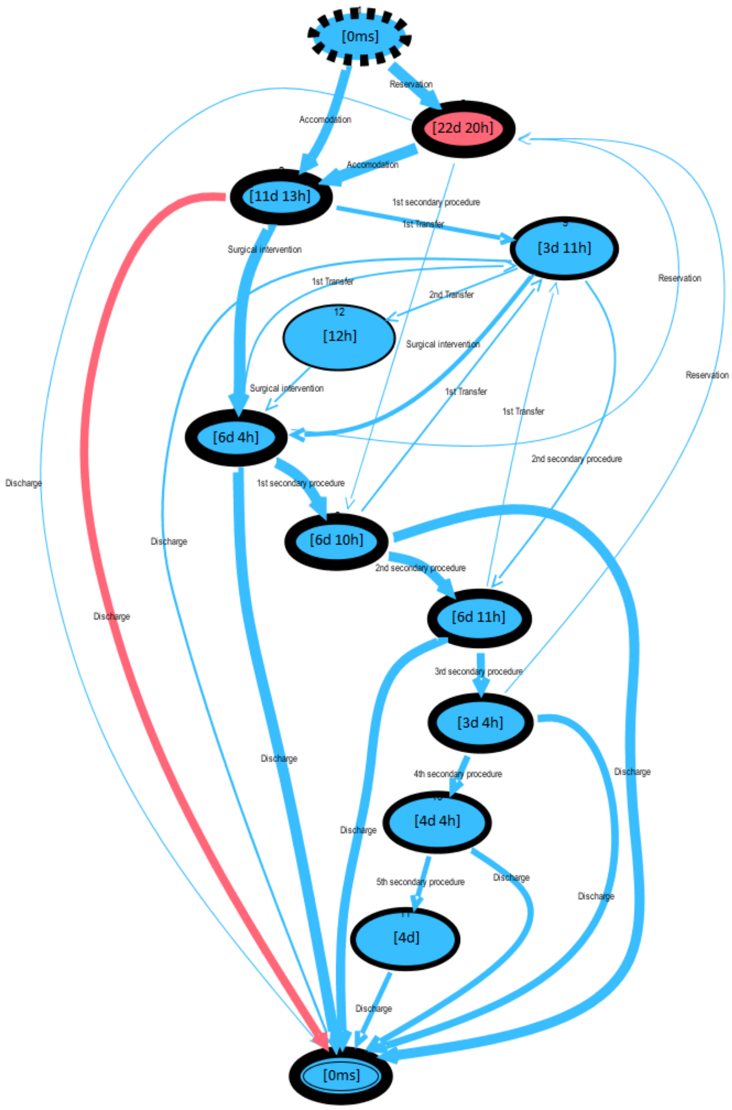


Fig. 36 Transition system (Totality of Patients)

We can see that the most of time, from the *sojourn time* point of view, is spent on the activity '*reservation*' where the patient is waiting for the accomodation to be hospitalized. Also the transition from the accomodation state to the discharge state takes a very long time. The standard deviation of these different average times is very high, meaning that the patients are spread out over a wide range of time, therefore the mean is not significant at all. There are patients that behaves different from other patients. For this reason we have split in different clusters the set of patients in order to discover interesting properties. Since, the primary of a hospitalization ward has the goal to minimize the time of the overall hospitalization procedure of a patient, we wanted to understand the bottleneck activities which slow down the entire hospitalization procedure. Therefore, we have performed an ad-hoc analysis for the patients who performed only the main surgical intervention without any secondary procedure and transfers to other hospitalization wards (in order to make uniform the data for the analysis). Since, we're interested in three particular hospitalization wards we have defined three different cluster of patients:

- patients hospitalized in the hospitalization ward '*0901*' (chirurgia generale)
- patients hospitalized in the hospitalization ward '*2601*' (medicina generale)
- patients hospitalized in the hospitalization ward '*3601*' (ortopedia e traumatologia)

for each cluster we have looked on how the patients are distributed over the time and we selected a 5% of patients with the lowest time duration of hospitalization procedure and a 5% of patients with the highest time duration ofhospitalization procedure. From this two subclusters we discovered process models, social networks and transition systems (looking only at the sojourn time point of view).
We started the analysis with the cluster of patients hospitalized in the hospitalization ward chirurgia generale. The process model obtained is:

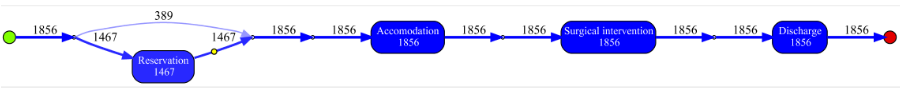


Fig. 37 Chirurgia generale

From the picture, points out, the most of patients ask for a reservation instead of being directly accomodated. Regarding the social network, we got trivially, a social network made up by only one hospitalization ward, since we're considering only patients of '*0901*' (chirurgia generale):

Fig. 38 Chirurgia generale

The derived transition system (sojourn time point of view) is:

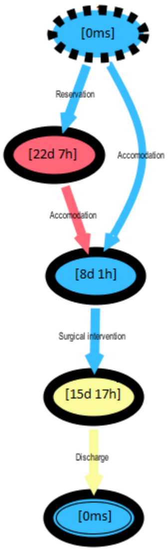


Fig. 39 Chirurgia generale

From the transition system points out that the patients take a very long time both to be accomodated and to be discharged. Then, we split the cluster chirurgia generale in two subclusters and we compared them between each other. For doing that, we first looked at the distribution of patients belonging to that cluster:

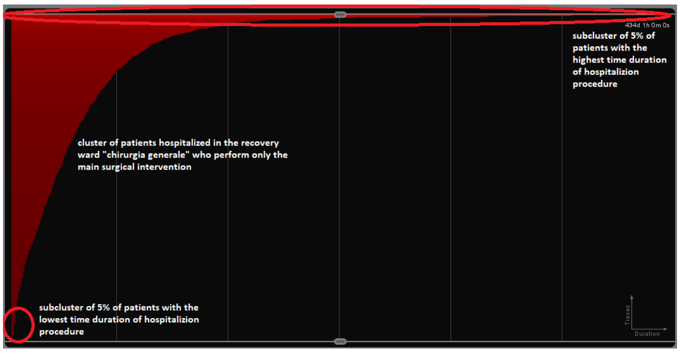


Fig. 40 Distribution of patients based on time duration

The Y-axis represents the patients and the X-axis represent the time duration. The 5% of patients with the lowest time duration of hospitalization procedure is lozalized in the bottom zone in red, instead the 5% of patients with the highest time duration of hospitalization procedure is lozalized in the top zone in red. The process model obtained from the subclusters are the following:

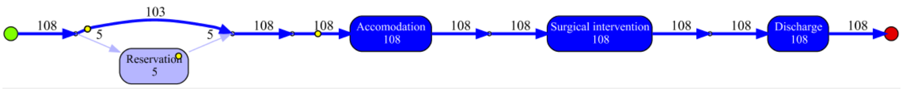


Fig. 41 5% of patients with the lowest time duration of hospitalization (Chirurgia generale)

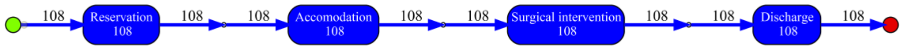
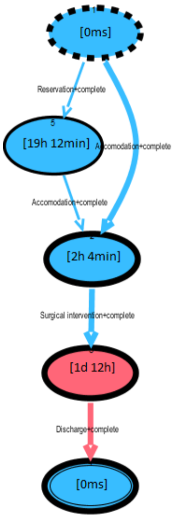


Fig. 42 5% of patients with the highest time duration of hospitalization (Chirurgia generale)

The difference that points out is on the most frequent path taken by the patients belonging to the two different subclusters. All the 5% of patients with the highest time duration of hospitalization (chirurgia generale) ask for a reservation, instead, the 5% of patients with the slowest time duration of hospitalization (chirurgia generale) are directly accomodated for hospitalization. The social networks obtained are the same as the one obtained for the main cluster chirurgia generale and for this reason we won't show them. Instead, it's important to show the transition systems, on the sojourn time point of view:



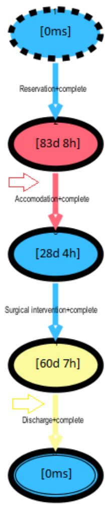


Fig. 43 5% of patients with the lowest time duration of hospitalization (Chirurgia generale)

Fig. 44 5% of patients with the highest time duration of hospitalization (Chirurgia generale)

The bottleneck is on the time that a patient is waiting for the accomodation (look at the red arrow). Moreover, also the time spent from the surgical intervention to the discharge of the patient is high (look at the yellow arrow). For the subclusters derived from chirurgia generale, we concluded:

- the hospitalization procedure is quick when the patients didn't ask for a reservation and they are directly accomodated for being hospitalized.
- the hospitalization procedure is mainly slowed down by the waiting time of the patients asking for the reservation but also on the time difference between the surgical intervention and the discharge of the patient (the primary of a hospitalization ward is more interested to minimize this time).

The procedure for the cluster medicina generale is impossible to be applied due to the small number of patients who performed only the main surgical intervention without transfers and secondary procedures (59 case IDs). On the other hand, we were able to analyze the cluster ortopedia e traumatologia. The process model obtained is:

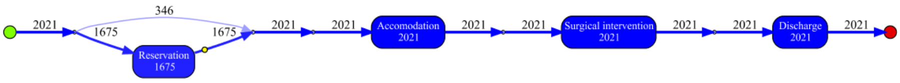


Fig. 45 Ortopedia e traumatologia

From the picture, points out, the majority of patients belonging to ortopedia e traumatologia asks for a reservetion in order to be accomodated for the hospitalization instead of being directly accomodated. Regarding the social network, we got trivially, a social network made up by only one hospitalization wards, since we're considering only patients of '*3601*' (ortopedia e traumatologia):

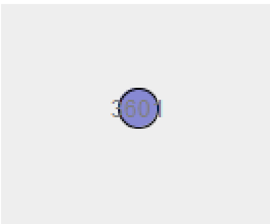


Fig. 46 Ortopedia e traumatologia

The derived transition system (on the sojourn time point of view) is:

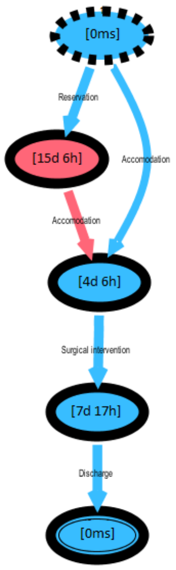


Fig. 47 Ortopedia e traumatologia

Then we have split the cluster ortopedia e traumatologia in two subclusters and we compared them between each other. The distribution of patients belonging to that cluster is the following:

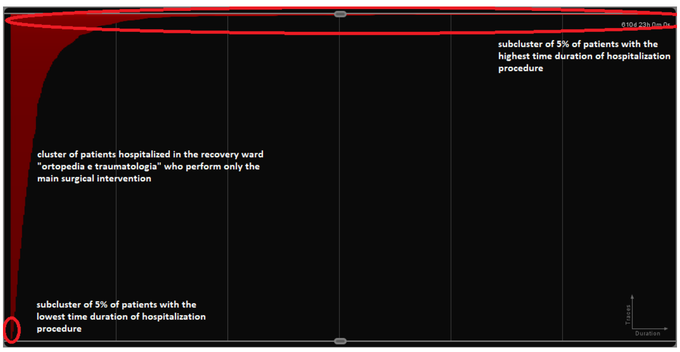


Fig. 48 Distribution of patients based on time duration

The Y-axis represents the patients and the X-axis represent the time duration. The 5% of patients with the lowest time duration of hospitalization procedure is lozalized in the bottom zone in red, instead the 5% of patients with the highest time duration of hospitalization procedure is lozalized in the top zone in red.

The process models obtained are the following:

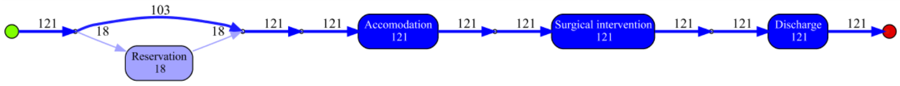


Fig. 49 5% of patients with the lowest time duration of hospitalization (Ortopedia e traumatologia)

Fig. 50 5% of patients with the highest time duration of hospitalization (Ortopedia e traumatologia)

The difference that points out is on the most frequent path taken by the patients belonging to the two different subclusters. The overall 5% of patients with the highest time duration of hospitalization (ortopedia e traumatologia) ask for a reservation, instead, the 5% of patients with the slowest time duration of hospitalization (ortopedia e traumatologia) is directly accomodated for hospitalization. The social networks obtained are the same as the one obtained for the main cluster ortopedia e traumatologia and for this reason we won't show them. Instead, it's important to show the transition systems, on the sojourn time point of view:

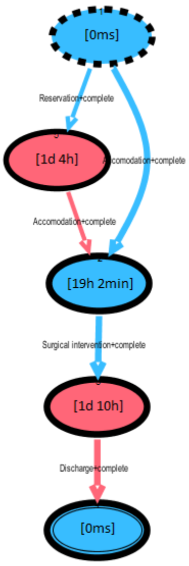


Fig. 51 5% of patients with the lowest time duration of hospitalization (Ortopedia e traumatologia)

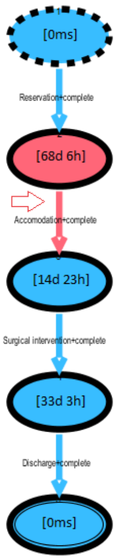


Fig. 52 5% of patients with the highest time duration of hospitalization (Ortopedia e traumatologia)

The bottleneck is on the time that a patient is waiting for the accomodation (look at the red arrow). Therefore, for the subclusters derived from ortopedia e traumatologia,

we concluded that the hospitalization procedure is mainly slowed down by the waiting time of the patients asking for the reservation of the hospitalization.
After did this in-depth analysis, we considered the set of clusters: [1], [2], [3], [4]. The process models obtained for the clusters [1] and [2] are:

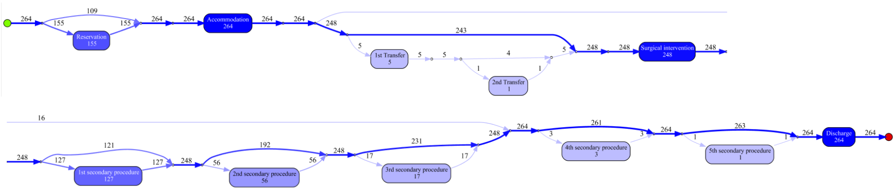


Fig. 53 Clinical path (Foreign patients)

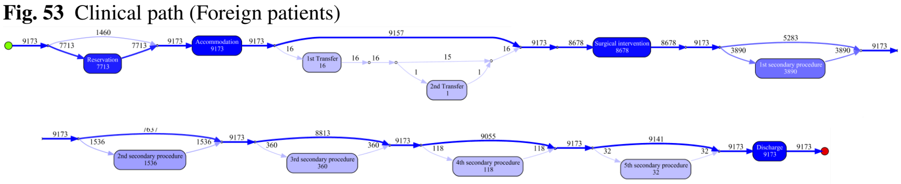


Fig. 54 Clinical path (Italian patients)

Both the process models are identical to the blueprint (in terms of behaviour of the patients), the only difference is on the number of process executions (264 for [1] and 9173 for [2]). Furthermore, looking at both the figures, we can see that the majority of both foreign and italian patients ask for a reservation, then they are accomodated for being hospitalized and after the surgical intervention they are discharged. On the other hand, the derived social network for both the clusters of foreign [1] and italian patients [2] are:

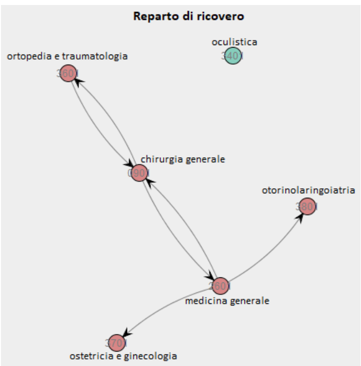


Fig. 55 Social network (Foreign patients)

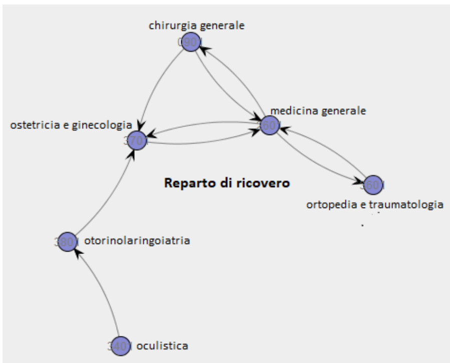


Fig. 56 Social network (Italian patients)

From the social network of the cluster of foreign patients [1] we can see that the hospitalization ward '*3401*' is not involved in any social relations, meaning that the patients hospitalized in '*3401*' (oculistica) won't be transferred in any other hospitalization wards. Instead, the social network of the cluster of italian patients [2] is more or less equal to the social network of the blueprint: it contains the same number of resources but the social relations are a subset.
The transition systems, on the *sojourn time* point of view, are:

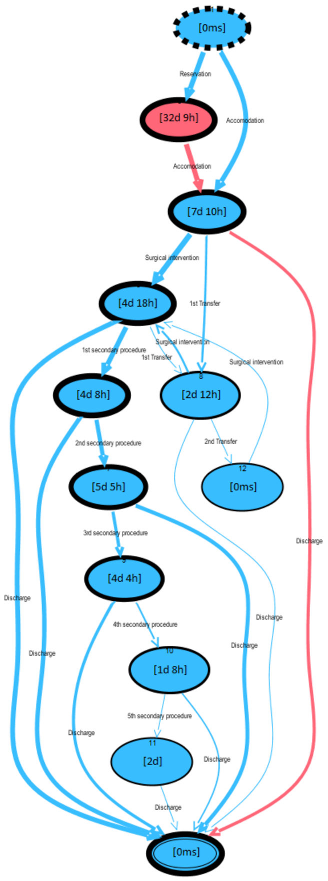


Fig. 57 Transition system (Foreign patients)

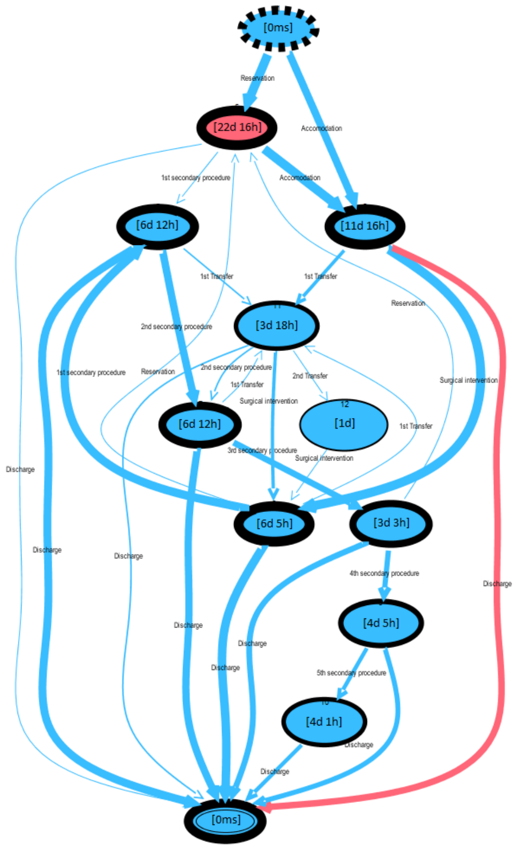


Fig. 58 Transition system (Italian patients)

Looking at both the transition systems we can see the bottleneck is on the activity '*reservation*'. We can see also that states and arcs associated with the cluster of italian patients are more thick than the states and arcs associated with the cluster of foreign patients, since the number of foreign patients is strictly lower than the number of italian patients.

At last, we have repeated the same steps for both the clusters [3], [4]. The derived process models are:

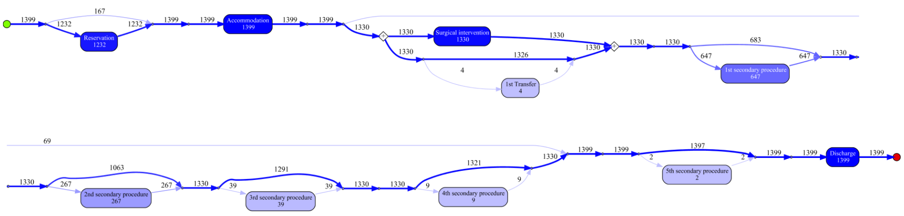


Fig. 59 Clinical path (≤25 years)

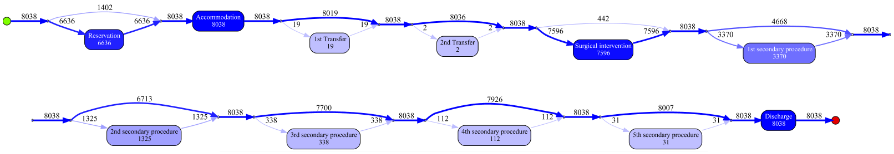


Fig. 60 Clinical path (>25 years)

The process models are similar but there are slightly differences: patients belonging to the cluster [3] performs only the '*1st transfer*' and not the '*2nd transfer*' . Note that the process model for [4] presents deviations regarding transfers, since they can happen also after the surgical intervention. Loooking at the process executions (1399 for [3] and 8038 for [4]) we can see that the number of patients belonging to [3] are strictly lower than the number of patients belonging to [4].
The derived social network for both the clusters [3] and [4] are:

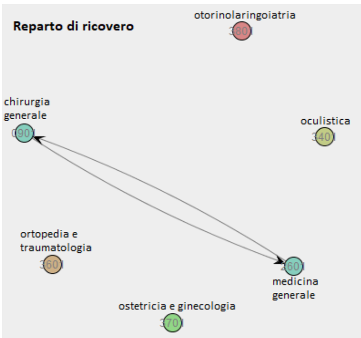


Fig. 61 Social network (≤25 years)

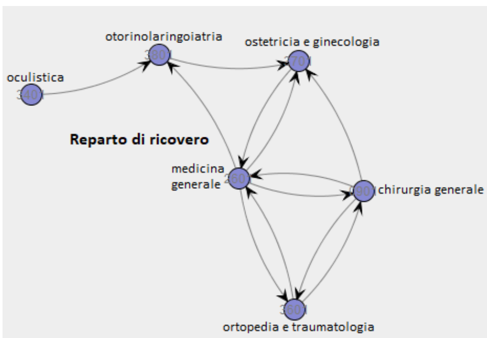


Fig. 62 Social network (>25 years)

From the social network of patients belonging to the cluster [3] the resources involved in social relations are both '*0901*' (chirurgia generale) and '*2601*' (medicina generale), meaning that patients are transferred from the hospitalization ward '*0901*' to '*2601*' and viceversa. Instead, the social network of patients belonging to the cluster [4] is equal to the social network of the blueprint. The transition systems, on the *sojourn time* point of view, are:

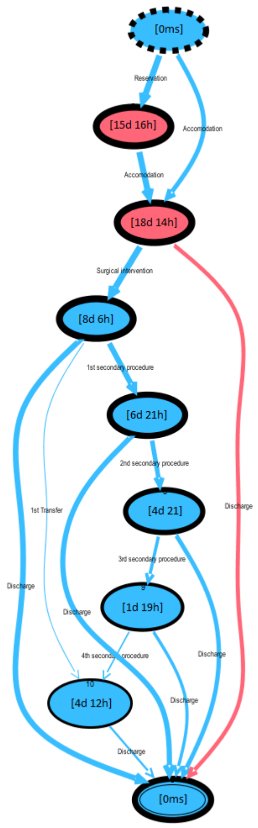


Fig. 63 Transition system (≤25 years)

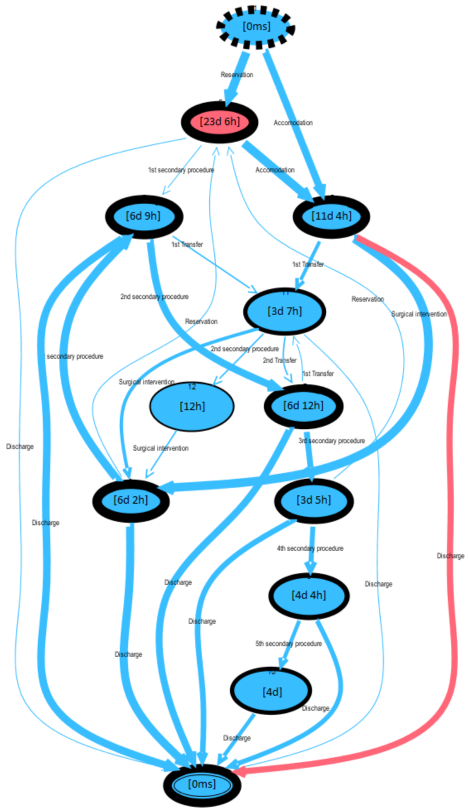


Fig. 64 Transition system (>25 years)

Looking at both the transition systems we can see the bottleneck is on the activities '*reservation*' and '*accomodation*' (only for patients belonging to [3]). We can see also that states and arcs associated with the cluster [4] are more thick than the states and arcs associated with the cluster of [3], since the number of patients with an age lower or equal than 25 years (resident in Latium) is strictly lower than the number of patients with an age greater than 25 years (resident in Latium).

5 Conclusions

In this paper, we have focused on the applicability of process mining in the healthcare domain. We have focused on the analysis of the results obtained for our case study. In particular, we have used the raw datasets of '*San Carlo di Nancy*' hospital in order to discover several non-trivial care processes, social networks and transition systems by using the ProM framework. Thus, we have shown that is possible to mine complex hospital processes giving insights into the process, that is, with existing techniques we were able to derive an understandable model for all the patients. In particular, we focused on obtaining insights into the careflow by looking at the control flow perspective and presenting the results showing the differences between the process models of the different cluster of patients. We have also look on the organizational perspective of the process models describing how the resources can communicate among them. In conclusion, we have also looked on the performance perspective of process models, analyzing the transition systems obtained from relevant event logs.

References

Aalst et al., 2009. Aalst, Wil M. P., Boudewijn F. van Dongen, Christian W. Gnther, Anne Rozinat, Eric Verbeek, & A Weijters 2009. ProM: The Process Mining Toolkit.

Aalst et al., 2018. Aalst, Wil M. P., Vladimir Rubin, B F Van Dongen, E Kindler, & C W Gnther 2018. Process mining: A two-step approach using transition systems and regions.

Buijs, 2010. Buijs, Joos 2010. Mapping Data Sources to XES in a Generic Way. Master's thesis, Eindhoven University of Technology.

Buijs, 2017. Buijs, Joos 2017 (accessed December 29, 2017). Introduction to Process Mining with ProM.

Buijs et al., 2017. Buijs, Joos, et al. 2017 (accessed December 29, 2017). Process Mining in Healthcare.

Buijs et al., 2012. Buijs, Joos C. A. M., Boudewijn F. van Dongen, & Wil M. P. van der Aalst 2012. On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery, pages 305–322. Springer Berlin Heidelberg, Berlin, Heidelberg.

Buijs et al., 2014. Buijs, J. C. A. M., B. F. van Dongen, & W. M. P. van der Aalst 2014. Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity. International Journal of Cooperative Information Systems, 23(01):1440001.

Cardoso et al., 2006. Cardoso, J., J. Mendling, G. Neumann, & H. A. Reijers 2006. A Discourse on Complexity of Process Models, pages 117–128. Springer Berlin Heidelberg, Berlin, Heidelberg.

Gruhn & Laue, . Gruhn, Volker, & Ralf Laue. Complexity metrics for business process models. In in: W. Abramowicz, H.C. Mayr (Eds.), 9th International Conference on Business Information Systems (BIS 2006), Lecture Notes in Informatics, pages 1–12.

Leemans et al., 2014. Leemans, Sander J. J., Dirk Fahland, & Wil M. P. van der Aalst 2014. Discovering Block-Structured Process Models from Event Logs Containing Infrequent Behaviour, pages 66–78. Springer International Publishing, Cham.

Leemans et al., 2015. Leemans, Sander J. J., Dirk Fahland, & Wil M. P. van der Aalst 2015. Exploring Processes and Deviations, pages 304–316. Springer International Publishing, Cham.

Leemans et al., 2016. Leemans, Sander J. J., Dirk Fahland, & Wil M. P. van der Aalst 2016. Using Life Cycle Information in Process Discovery, pages 204–217. Springer International Publishing, Cham.

Mans et al., 2009. Mans, R. S., M. H. Schonenberg, M. Song, W. M. P. van der Aalst, & P. J. M. Bakker 2009. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital, pages 425–438. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rojas et al., 2016. Rojas, Eric, Jorge Munoz-Gama, Marcos Seplveda, & Daniel Capurro 2016. Process mining in healthcare: A literature review. Journal of Biomedical Informatics, 61(Supplement C):224 – 236.

Rozinat & van der Aalst, 2008. Rozinat, A., & W.M.P. van der Aalst 2008. Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1):64 – 95.

Russo & Mecella, 2013. Russo, Alessandro, & Massimo Mecella 2013. On the evolution of process-oriented approaches for healthcare workflows. International Journal of Business Process Integration and Management, 6(3):224–246. PMID: 56962.

Song & Aalst, 2007. Song, Minseok, & Wil M. P. Aalst 2007. Supporting Process Mining by Showing Events at a Glance.

van der Aalst et al., 2011. van der Aalst, W.M.P., M.H. Schonenberg, & M. Song 2011. Time prediction based on process mining. Information Systems, 36(2):450 – 475. Special Issue: Semantic Integration of Data, Multimedia, and Services.

van der Aalst et al., 2003. van der Aalst, W.M.P., A.J.M.M. Weijter, & L. Maruster 2003. Workflow Mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering, 16:2004.

van der Aalst, 2011. van der Aalst, Wil M. P. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer Publishing Company, Incorporated.

van der Aalst et al., 2010. van der Aalst, Wil M. P., Maja Pesic, & Minseok Song 2010. Beyond Process Mining: From the Past to Present and Future, pages 38–52. Springer Berlin Heidelberg, Berlin, Heidelberg.

van der Aalst et al., 2005. van der Aalst, Wil M. P., Hajo A. Reijers, & Minseok Song 2005. Discovering Social Networks from Event Logs. Computer Supported Cooperative Work (CSCW), 14(6):549–593.

van der Aalst & Song, 2004. van der Aalst, Wil M. P., & Minseok Song 2004. Mining Social Networks: Uncovering Interaction Patterns in Business Processes, pages 244–260. Springer Berlin Heidelberg, Berlin, Heidelberg.

van der Aalst & van Dongen, 2013. van der Aalst, Wil M. P., & Boudewijn F. van Dongen 2013. Discovering Petri Nets from Event Logs, pages 372–422. Springer Berlin Heidelberg, Berlin, Heidelberg.

van Dongen et al., 2016. van Dongen, B. F., J. Carmona, & T. Chatain 2016. A Unified Approach for Measuring Precision and Generalization Based on Anti-alignments, pages 39–56. Springer International Publishing, Cham.

Verbeek et al., 2011. Verbeek, H. M. W., Joos C. A. M. Buijs, Boudewijn F. van Dongen, & Wil M. P. van der Aalst 2011. XES, XESame, and ProM 6, pages 60–75. Springer Berlin Heidelberg, Berlin, Heidelberg.

Verbeek, 2013. Verbeek, H. M. W. (Eric) 2013. BPI Challenge 2012: The Transition System Case, pages 225–226. Springer Berlin Heidelberg, Berlin, Heidelberg.

Weijters et al., 2006. Weijters, A, Wil M. P. Aalst, & Alves A K Medeiros 2006. Process Mining with the Heuristics Miner-algorithm.