Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Corso di Laurea Magistrale in Ingegneria Matematica

Monitoring and Modeling hospital networks using administrative data

Relatore:
Prof. Anna Paganoni

Co-Relatore:
Dott. Francesca Ieva

Tesi di Laurea Magistrale di:
Alessandra Grossi
Matr. 799259

Anno Accademico 2013-2014

Contents

List of Figures

List of Tables

Abstract

This work arises within the FARB project “Public Management Research: Health and Education System Assessment”. The main interest of the project lies in improving the quality of health and educational services. In this work we use administrative data on hospitalizations due to chronic cardiovascular diseases, for assessing healthcare quality in Lombardia (a northern Italy regional disctrict whose capital is Milan).
The topic of this thesis is the analysis of data coming from hospital discharge papers of patients resident in Lombardia, which belong to the BDA (Banca Dati Assistito, i.e., the database of individuals registered within the Lombardia healthcare system). The aim is firstly to outline hospitals that had an unusual behaviour with respect to mortality and re-hospitalization, then to summaryze the techniques to assess quality on the run and lastly to analyze hospitals in Lombardia as a network.
The structure of the work is as follows:

- In Chapter 1 we present the motivating problem and the dataset.
- In Chapter 2 and Chapter 3 we present Cumulative Sum Charts and funnel plots and their use.
- In Chapter 4 evaluation of hospital performances in terms of mortality rate.
- In Chapter 5 evaluation of hospital performances in terms of re-hospedalization rate.
- In Chapter 6: we present the network approach.
- In Chapter 7: we introduce the hospital networks.
- In Chapter 8: we present the gravity model and the exponential random graph model for the analysis of network data, applying them to the case study of interest .

All the analysis have been carried out using the statistical software R.3.0.1.

Keywords: Funnel plots, Hospital Monitoring, Heart Failure, Networks, ERGM.

Sommario

Questa tesi nasce all'interno del progetto FARB "Public Management Research: Health and Education System Assessment", il cui scopo è volto al miglioramento dei sistemi scolastico e sanitario. In questo elaborato introduciamo una valutazione delle strutture ospedaliere, utilizzando i dati raccolti nella Banca Dati Assistito, un dataset amministrativo che tiene conto dei dati relativi alle ammissioni ospedaliere dei pazienti residenti il Lombardia. In questa tesi ci concentriamo sulle informazioni riguardanti pazienti affetti da una patologia cardiaca cronica, lo scompenso cardiaco e usiamo i dati ottenuti dalle Schede di Dimissione Ospedaliera per valutare le strutture ospedaliere.
Un primo passo nell'inquadramento del problema consiste nel presentare ed applicare sul dataset fornitoci metodi già affermati per la valutazione degli ospedali. Usando i funnel plot caratterizziamo il comportamento delle strutture nei riguardi della percentuale di decessi e di riospedalizzazioni, evidenziando eventuali strutture che presentano comportamenti estremi, sia in positivo che in negativo. In un secondo momento introduciamo altre tecniche utilizzate per la valutazione e infine proponiamo un approcio di tipo Network per studiare la rete di ospedali.
La struttura è come segue:

- Nel Capitolo 1 presentiamo il problema, gli scopi e il dataset.
- Nei Capitoli 2 e 3 presentiamo le Cumulative Sum charts, i funnel plot e il loro utilizzo nella valutazione delle strutture.
- Nel Capitolo 4 valutiamo le strutture ospedaliere utilizzando come output la percentuale di decessi.
- Nel Capitolo 5 valutiamo le strutture ospedaliere utilizzando come output la percentuale di riospedalizzazioni.
- Nel Capitolo 6 presentiamo l'approccio network.
- Nel Capitolo 7 presentiamo e analizziamo diversi network di ospedali.
- Nel Capitolo 8 presentiamo vari modelli generalizzati che utilizziamo nell'analisi dei network: diversi esempi di "gravity model" and "exponential random graph model", che applichiamo al caso di studio.

Tutte le analisi sono state effettuate utilizzando il software R.3.0.1.

Introduzione

I dati amministrativi, ovvero le informazioni raccolte e conservate da istituzioni pubbliche, sono sempre più usate in campi di valutazione e di monitoraggio di strutture. Nel campo ospedaliero, diversi autori [5, 15] suggeriscono di usare queste risorse per ottenere indici clinici, per valutare le prestazioni degli ospedali e per ricerche epidemiologiche [15, 18]. Un'altra possibilità già esplorata [4] è quella di usare dati amministrativi, che sono continuamente aggiornati, per poter sorvegliare la performance delle strutture e accorgersi rapidamente di possibili deterioramenti [31, 5]

I vantaggi che derivano dall'utilizzo di queste banche dati sono certamente grandi: la dimensione campionaria è estremamente elevata, arrivando spesso a coprire la totalità della popolazione disponibile, i costi di raccolta sono nulli, essendo già disponibili, il tempo di osservazione è estremamente lungo e infine, grazie all'unicità dei codici identificativi delle persone, è possibile collegare diverse banche dati e arricchire i dataset. Anche se i dati clinici sono spesso ridotti e lacunosi, poiché lo scopo della raccolta non era né la valutazione delle strutture, né possibili analisi epidemiologiche, l'utilizzo di queste risorse è prezioso per la valutazione delle strutture ospedaliere.

Il dataset esaminato comprende i dati provenienti dalle Schede di Dimissione Ospedaliere (SDO) di tutti i residenti della Lombardia, di cui sono stati selezionati solo quelli ricoverati per scompenso cardiaco tra il 2000 e il 2012. Le informazioni, che provengono dalle SDO, sono raccolte nella Banca Dati assistito, la banca dati che tiene traccia degli eventi medici: per ogni evento di scompenso cardiaco possiamo quindi ricavare informazioni sui trattamenti ricevuti dai pazienti. La scelta dei pazienti soggetti alla patologia di interesse è una scelta delicata, poiché la diagnosi per scompenso cardiaco comprende diversi codici ed è complessa da definire in un dataset amministrativo. La selezione dei pazienti trattatati in questa tesi viene quindi descritta in [18]. Dopo esserci occupati della pulizia del dataset, introdurremo metodi per l'identificazione di quelle strutture che presentano comportamenti estremi riguardo alla percentuale di decessi e a quella di riospedalizzazioni e spiegheremo possibili tecniche per un continuo monitoraggio delle strutture. Useremo infine questi strumenti per individuare effettivamente le strutture che presentano un comportamento anomalo. In una seconda parte della tesi introdurremo un'analisi di rete. Gli ospedali saranno quindi collegati gli uni con gli altri e i loro rapporti saranno definiti sia dal flusso di pazienti che si muove nella rete degli ospedali, sia dalla somiglianza tra strutture. Dopo aver presentato questi possibili network di ospedali, presenteremo dei GLM adattati ai network introdotti.

Tutte le analisi sono state condotte usando il software statistico R.3.0.1[28].

Part I
Introduction and Motivations

Introduction

Administrative data, collected mainly for administrative purposes, are increasingly being used to derive hospital-specific clinical indicators and to evaluate their performance [5, 15]. Bottle and al. [4] use administrative data to help hospitals in identifying areas in need of self-monitoring or to identify structures that present a too high mortality rate [31, 5]. The use of administrative data is already well established in epidemiological researches [15, 18]. Advantages in their use can be identified in large sample size, no gathering costs, really long observational period and sometimes the possibility of linking new databases information, thanks to the uniqueness of encrypted ID codes. Problems and drawbacks are clearly present: data are collected for different goals than those needed, causing lack of informations. Despite the possible drawbacks, administrative data are a powerful resource that we must use to the fullest.

Our dataset comes from dimission hospital discharge papers (Schede di Dimissione Ospedaliera, SDOs) of patients admitted for heart failure events [18]. In Lombardia (a northern Italy regional disctrict), all the SDOs and other data related to services obtained from medical events, are collected in several databases. These databases are organized in a patient-centerd database (BDA, Banco Dati Assistito), so that is possible to link drugs, procedures, etc. to patients and identify their medical history.

Our interest lies in the hospitals capability of treating patients affected by Heart Failure (HF). The choice of criteria to select patient affected by a specific disease is one of the most relevant issues when dealing with administrative data. The patients selection is detailed in [18]. Fot the aims of this work we select the subsect of all the patients that were identified as sure HF patients.

In conclusion the database we have is a collection of HF events, with every row of the dataset regarding one event for a patient. A row includes all the information we have about that Hearth Failure event.

The aim is to outline hospitals that had an unusual behaviour with respect to mortality and re-hospitalization, then to summaryze the techniques to assess quality on the run and lastly to analyze hospitals in Lombardia as a network.

As a first step we explore the data, checking for coherency, missing ID's and other similar problems, then we select the subset of interest. Then we present two well established approaches to evaluate hospitals: the construction of funnel plot for identifying outliers and creation of CUSUM tables for monitoring performance over time[4]. We explain advantages and drawback and we use funnel plot outlier detection on our data.

In the second part of the work we propose to study different hospitals networks. We propose a flow network, where hospitals are connected by patients flows and a similarty network, where similar structures are considered near. We analyze networks' features and we model all the hospital networks we propose.

All the analysis have been carried out using the statistical software R.3.0.1[28].

1 Data

In this section we will present the dataset and the preprocessing neded for introducing both monitoring and modelling techniques described in the following chapters.

1.1 Dataset introduction

The original dataset, that from now on will be referred to as the Complete Dataset, is an administrative dataset containing all the data from the hospital discharge papers of patients resident in Lombardia that are admitted for Heart Failure (HF) events. The hospitalizations take place from 2000 to 2012 mainly, but not only, in Lombardia. Since the dataset is centered on patients that are resident in Lombardia, but records all data regarding them, also events that are related to admission in other regions are present in the dataset.
In the Complete Dataset patients are divided in four groups, one of which is composed by events related to sure HF as explained in [18]. This subset is the one we select for further analysis.
As we see in Table 1 the dataset is event based: a row represents a HF event for one patient and any patient must have at least one Heart Failure event to enter the study. Every event, patient and hospital has an encrypted code, so we can group events for patient or hospital.

Row	Patient ID	Event ID	Age	Date of admission	...
1	10,000,024	1	61	10/09/2000	...
2	10,000,024	2	62	11/10/2001	...
3	10,000,032	3	45	05/04/2006	...
4	10,000,048	4	70	...	...

Table 1: Example of data from BDA. Any row represents a single Heart Failure event. We see that a patient can have more than one HF event during the considered period.

The events recorded in the original dataset are 701,701 corresponding to 371,766 patients admitted into 1,035 hospitals. Among clinical informations there are: sex, age, date of admission and discharge, length of stay, admission ward, date of death (collected from the civil registry), AHRQ and HCC80 codes (that were used to identify the group of sure HF events from the Complete Dataset), a flag indicating if the patient dies during the hospitalization, flags for comorbidities, a flag indicating if the patient visited Intensive Therapy (IT) during the hospitalization, a flag indicating if the recovery is a rehabilitation one, a flag indicating the presence of an event of shock and flags indicating the use of cardiochirurgy, a defibrillator implant, and other procedures. In the evaluation of comorbidities we use a risk index proposed in [13]. This index mixes the comorbidity flags; high values of the index are associated to high risk patients, low values to low risk patients. Another variable that is available is the total espense of admission, that can be used to compare hospitals regarding to cost-effectiveness.
During a HF event the patient can be transferred from an hospital to another. When patients are discharged and readmitted to another hospital in the same day they are said to be transferred. About 12% of events in the dataset are events with transfers. In the dataset all the transfers are grouped as part of the same event and the number of transfers in an event are identified in a column counting the number of different hospitalizations that occurred in the same event.
With more than one hospitalization in a single event, Length Of Stay (LOS) and all the procedures are cumulative and the hospital label is the hospital of first admission. So if a patient is firstly admitted in an hospital without intensive care, but then is transferred somewhere else, the event can have a positive IT flag. Also, if the patient dies in the last hospital, the death is

marked under the first hospital. This assignation problem is widely discussed in literature [5, 4] and there isn't a simple or unique solution, but in the database we deal with events that are already grouped. As such we use the cumulative events we have without discussing it further.

1.2 Data quality control and subset selection.

From the Complete Dataset of 701, 701 events we outline 489, 903 that were assigned in the group of sure HF events. We then discard patients that presented invalid or missing identifiers, hospital of admission and date of admission/discharge. Checking date coherency we remove also all the events that have a date of death preceding the date of admission. Then we check coherency between flag for intra-hospital death and the dates of admission, discharge and death. A great number of patients (more than 1, 500) have positive flag of intra-hospital death and a date of death following the date of dimission, so that they are marked as dead in the hospital and have a date of death posterior to the dimission date, or vice versa people that have negative intra-death flag, but have a date of death included between date of admission and discharge. Specialists say to consider the flags of intra hospital mortality as the reliable ones. Indeed, an intra-death hospital flag is collected on the SDO and death report errors are very unusual. Dates of death are recovered from civil registry and are not always accurate (in most of the cases data of discharge and data of death differs only for one or two days).
Finally we discard also underage patients, we choose not to deal with them, because they behave differently from the typical HF patient.
We have now 488,820 events from 254,852 patients that are admitted to 939 hospitals. Among these we retained only hospitals with more than 20 patients each, in order to have proper dimensions when carrying out estimates in the analysis. This reduced the dataset to patients admitted to 275 hospitals. The consistent reduction is due to the fact that the most part of the hospitals are very small as we can see in Figure 1.

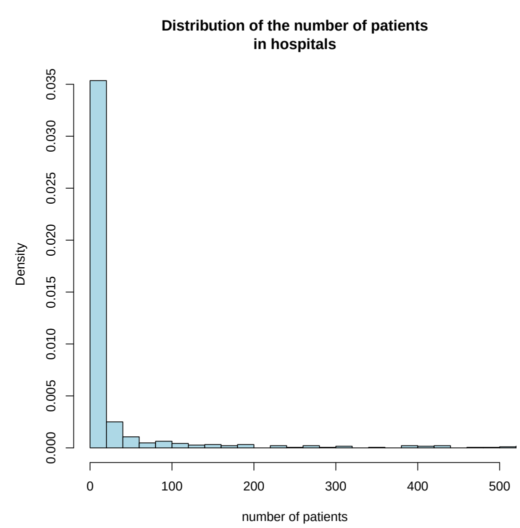


Figure 1: Histogram of number of patients per hospital. To show the high frequency of small hospitals only structures with less than 500 patients are shown.

A possible explanation of the high number of small hospitals is that events recorded include also recovers outside Lombardia, probably these structures gather less than 20 patients.

1.3 Dataset description

In Table 2 we explain the meaning of the variables we use in the study. They are already at hand in the original dataset with the exception of the re-hospitalization flag that is extrapolated from the data. All flag variables are 1 if the named event happens, 0 if not.

Variable	Description
Dead	Intra-hospital death flag
IT	Flag for presence of Intensive Theraphy during the event
Cardiology	Flag for the use of Cardiochirurgy
ICD	Flag for a Defibrillator Implant
DRG	Flag that explain if the admission was labeled as a surgery one (administrative choice)
Shock	Flag for presence of a shock event
PTCA	Flag for Percutaneous Transluminal Coronary Angioplasty
CABG	Flag for Coronary Artery Bypass Graft
MDC5	Flag that explain if the admission is labeled as a Major Diagnostic Category 5
Risk	Risk index obtained from comorbidities flags goes from -2 (very low risk) to 13 (very high risk)
LOS	Length Of Stay goes from 0 to 958
Age	Age of the patient at the begining of the event goes from 18 to 108
ReH	Flag that specifies if the patient is re-admitted after the current event

Table 2: List and explanation of variables analyzed in the study.

In Tables 3 and 4 some summary statistics concerning the dataset are reported.

	Number	
Patients	$253,475$	
Hospitals	275	
Events	$485,693$	
	Females	Males
Patients	53%	47%

	percentage
Dead	0.086
ReH	0.400
IT	0.200
Cardiology	0.053
ICD	0.012
DRG	0.150
Shock	0.024
PTCA	0.054
CABG	0.030
MDC5	0.990

Table 3: Summary of the dataset with hospitals with more than 20 events each.

	min	max	median	mean	SD
Age (years)	18	108	79	77.000	11.06
Risk	-2	13	1	0.990	1.42
LOS (days)	0	958	10	14.000	14.29

Table 4: Summary statistics for the variables of interest reported in Table2. Only hospitals with more than 20 cases are considered.

About 88% of the events are without in-event transfers, while 99% of the events are made by 3 or less hospitalization. We decide to use all the data in the descriptive analysis, while in the modelling we reduce the dataset only to events without transfers, since in these cases there is a one to one corrispondence between patients and hospitals.
The distribution of patients that have respectively K events is another feature of interest. In Table 5 we can see the distribution of number of events per patient.

Number of events per patients	%	Cumulative %
1	60.2	60.2
2	19.9	80.1
3	8.8	88.9
4	4.5	93.5
≥5	6.5	100

Table 5: Percentage and cumulative percentage of number of HF events per patient.

Almost 60% of patients have only one event, while only 6.5% have more than 5 events. We see both from the table and from other analyses that the percentage of rehospitalization increases with the number of the hospitalization itself. For example while less than 40% of patients at their first event will have another hospitalization, among patients at their second event about 50% will be rehospitalized, and so on.

Part II
Monitoring of hospitals by means of funnel plots and cumulative sum charts

Quality control in medicine has gained more and more interest over recent years. Not only industrial concepts of quality control have been thought as useful in healthcare monitoring [33, 25], but also a standard tool within meta-analysis like the funnel plot is now used in hospital evaluation [30, 19]. It's clear by now that medical performance needs to be adequately monitored and in this chapter we introduce the use of funnel plots and cumulative sum charts to this aim.

2 An introduction to funnel plots

Institutional comparisons can be handled in different ways [1]. In literature such comparisons commonly lead to the production of comparison tables, in which institutions are ranked according to a chosen performance indicator. The problem is that tables lead to a rank ordering, that in general is very questionable.
The funnel plots introduced by Spiegelhalter[30] avoids spurious ranking and in the mean time provides a strong and visual indication of outlier performances. Advantages include the display of the variable of interest, a graphical check of the relationship between event rate and volume of cases, and allow increased variability among smaller institutions. Furthermore an intuitive choice of axes can make the funnel plot a useful way to explain data.
In Figure 2 we see an example of a funnel plot. As we can see in the given example, for every hospital is plotted the rate of re-hospitalization against the volume. On the plotted data are then superimposed 95% (2 standard deviation) and 99% (3 standard deviation) confidence limits around the overall re-hospitalization rate. This funnel plot find the majority of the institutions lying within the 99% limits, while it clearly show hospitals that have an outlier behaviour.

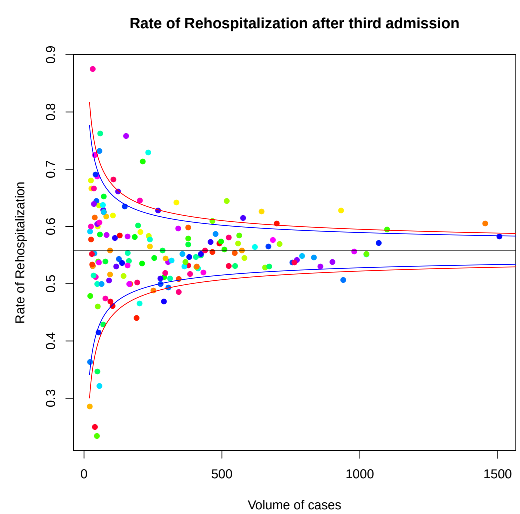


Figure 2: Example of a general funnel plot.

A funnel plot consists of four main components:

1. An indicator response Y (Rate of Re-hospitalization in Figure 2) on the Y-axis.
2. A target θ_0 for the response, representing a desired expectation i.e., $E(Y|\theta_0) = \theta_0$ when units are under control (θ_0 is around 0.56 in Figure 2).
3. A precision parameter ρ on the X-axis. ρ denotes the accuracy the indicator is measured by and is usually proportional to the inverse variance of the under control distribution: $\rho = g(\theta_0)/Var(Y|\theta_0)$, for some function g. The choice of ρ is arbitrary to some degree, but it is better to select an intepretable precision parameter, so g is chosen accordingly. When Y is a proportion (for example the rate of re-hospitalization in Figure 2) we have $V(Y|\theta_0) = \theta_0(1-\theta_0)/n$ and since the sample size is an interpretable measure of precison we set $\rho = n$ and $g(\theta_0) = \theta_0(1-\theta_0)$.
4. Control limits of level $1-p$, which are chosen so that the probability of going outside these limits for an under control unit is equal to p.

Given a series of I observations with indicators y_i and associated precisions ρ_i, a funnel plot consists of a plot of y_i against ρ_i, with target θ_0 shown by a horizontal line and control limits plotted as a function of ρ_i. Control limits are independent of the data being plotted.
For a more complete discussion about funnel plots we refer to [30].

2.1 Funnel limits

In many circumstances we can assume an approximate normal distribution:

$$Y|\theta_0 \sim N(\theta_0;\ g(\theta_0)/\rho) \tag{1}$$

In so doing the control limits will result in:

$$y_p(\theta_0; \rho) = \theta_0 + z_p\sqrt{g(\theta_0)/\rho} \tag{2}$$

where z_p is such that $\phi(z_p) = P(Z \leq z_p) = p$ for a standard normal distribution Z.

2.2 Overdispersion

If the target distribution doesn't express the variability of the under control units properly, a lot of institution will result outside the funnel even if they are not extreme. We refer to this saying they are overdispersed around the target. If this happens the appropriateness of the limits evaluated with 2 is questionable and proper verification for overdispersion may be needed.
An overdispersed behaviour is generally due to latent covariates: although each one may have a small impact on the outcome, when taken together they can lead to an excess variability among units that are under control. If this is not properly taken into account there will be an inappropriate number of units identified as special cases by the funnel plot. Since we build the limits at level 0.95 and 0.99, having half of the structures outside the funnel is a clear signal of overdispersion.
Two automatic procedures that expand the funnel limit are proposed in [30]. We shall now explore these two basic statistical models for over-dispersion: first the 'extreme-effects' formulation, which inflates the variance with a over-dispersion factor and then a 'random-effects' formulation which adds a constant term to the sampling variance of each unit.

Multiplicative (or random-effects) formulation

We assumed before that $\rho = g(\theta_0)/Var_0(Y|\theta_0)$, where the subscript on Var_0 has now been introduced to indicate the case of no over-dispersion.

The random effects formulation is a multiplicative approach that introduces an over-dispersion factor ϕ that inflates the null variance, so that the variance is now evaluated as:

$$Var(Y|\theta_0;\rho;\phi) = \phi \cdot Var_0(Y|\theta_0;\rho;\phi) = \frac{\phi g(\theta_0)}{\rho} \tag{3}$$

Suppose we have a sample of I units that we assume to be under control, then:

$$\hat{\phi} = \frac{1}{I}\sum_i \frac{(y_i - \theta_0)^2 \rho_i}{g(\theta)} = \frac{1}{I}\sum_i z_i^2 \tag{4}$$

Where z_i are called standardized Pearson residuals.

so the original funnel limits, defined in (2), are inflated by $\sqrt{\hat{\phi}}$. Over-dispersed control limits are now evaluated as:

$$y_p(\theta_0;\rho) = \theta_0 + z_p\sqrt{\hat{\phi} \cdot g(\theta_0)/\rho} \tag{5}$$

The problem to estimate ϕ in (4) lays in identifying the units under control to be used in the evaluation of ϕ. Sure enough, if out-of-control units are included in this estimation, they increase the estimate of ϕ, widen the funnel limits and include outliers inside the funnel. To avoid this particular issue we used the 'Winsorised' estimates, proposed in [30].

1. Choose a percentile q (and $1 - q$).
2. Evaluate the "naive Z-scores" such as: $z_i = \frac{y_i - \theta_0}{Var(Y|\theta)}$.
3. Rank istitution according to this Z-scores.
4. Identify z_q andz_{1-q}, the $100q\%$ most extreme top and bottom naive Z-scores.
5. Now set the lowest $100q$ % of Z-scores equal to z_q, and the highest $100q\%$ of Z-scores to z_{1-q}. Denote the resulting set of Z-scores as $z_i^W(q)$.

$\hat{\phi}$ may be estimated as:

$$\hat{\phi}^W = \frac{1}{I}\sum_i z_i^W(q)^2 \tag{6}$$

If there is no true over-dispersion, then $I\hat{\phi}^W$ has approximately a χ^2 distribution (see [30]), so $E(\hat{\phi}^W) = 1, \quad V(\hat{\phi}^W) = 2/I$.

Therefore we assume overdispersion only if

$$\hat{\phi}^W > 1 + 2 \cdot \sqrt{2/I} \tag{7}$$

If $\hat{\phi}^W < 1 + 2 \cdot \sqrt{2/I}$ we will set $\phi = 1$.

Additive (or outlier-effects) formulation

The second approach inflates funnel limits in an additive way [10]. This approach assumes that Y_i has expectation $E(Y_i) = \theta_i$, variance $V(Y_i) = s_i^2$, and that for under control units θ_i is distributed with mean θ_0 and standard deviation τ.

The parameter τ is estimated using the moments estimator:

$$\tau^2 = \frac{I\hat{\phi} - (I-1)}{\sum_i w_i - \sum_i w_i^2/\sum_i w_i} \tag{8}$$

Where $w_i = 1/s_i^2$ and $\hat{\phi}$ is as previously defined. If $\hat{\phi} < (I-1)/I$, then τ^2 is set to 0.

If $\hat{\phi} < (I-1)/I$ τ is set to 0.
The original funnel limits, defined in (2) are inflated and over-dispersed control limits are now evaluated as:

$$y_p(\theta_0; \rho) = \theta_0 + z_p \sqrt{g(\theta_0)/\rho + \tau^2} \tag{9}$$

3 Cumulative Sum chart

Every process can be seen as a wave-like motion: if we consider in-hopital mortality we see that patients that dies are never equally distributed in time. An hospital performance will be better or worse depending on various issues. These fluctuations around the mean can never be eliminated. An essential question is how to detect a degradation in a process. Statistical Process Control (SPC) consider the process variability as an index that identify if a process is under control or not. In theory, all points situated outside of the normal variation, indicate that the process is possibly out of control.

In healthcare literature, different kinds of SPCs are used. We thereby present Cumulative Sum charts, or CUSUM charts. These charts are based on sequential monitoring of the cumulative performance over a period of time [22]. The difference with respect to funnel plots is that this can be defined as an “on the run” approach. Indeed CUSUM charts are updated after each event so that there can be a real-time monitoring of performance, while more data are needed to evaluate a structure with funnel plots. Their strenght is that they identify subtle and slow degradation in a process and alert for this degradation as soon as possible [34, 4, 26].

Where their strength is in identify degradation in a single structure in a short amount of time, our data cover a very long period of time and our aim is to compare structures, so there is little use in the CUSUM approach. We shortly discuss the use and utility of the Cumulative failure charts in assessing hospital performance and the reasons why we won’t show CUSUMs analysis.

3.1 Cumulative failure charts

The CUSUM charts for monitoring hospital performance usually work on binary events, such as death/alive, or presence/absence of rehospitalization within a certain amount of days. In this example we choose the mortality as an outcome.

The cumulative failure chart (Figure 3) is the simplest form of a CUSUM chart. The chart is constructed by the cumulative number of events (failures), on the vertical axis, plotted against the total number of events on the horizontal axis. Each death makes the graph grow up, with a 100% event rate we have a 45° slope, while a horizontal line corrispond to the case of no events. Boundary lines to evaluate quality can be constructed. These boundaries, or alert lines, are costructed with accepted (p_0) and the unaccepted (p_1) failure rate. Values for the two alert lines in a failure chart are evaluated as in paper [25]. The two lines identify the acceptable failure rate and the unacceptable failure rate. If the graph of cumulative failures exceeds the upper boudary line (unacceptable failure rate), the failure rate is higher than the unacceptable rate (p_1). If the graph crosses the lower boundary line (acceptable failure rate), the failure rate is lower than the acceptable rate (p_0). An acceptable process shows a graph with a slope towards the lower boundary line.

Example of Cumulative failure charts

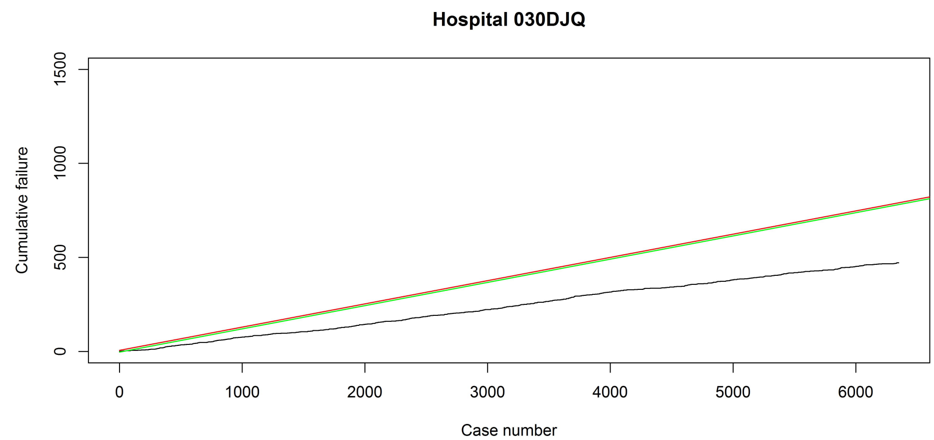


Figure 3: Cumulative failure chart for a generic hospital.

In Figure 3 we use the in-hospital mortality as a response (1 if the patient dies, and 0 if he/she remains alive). The p_0 and p_1 values are evaluated using the dataset, with p_0 beeing the global mean and p_1 beeing the global mean plus two standard deviation. The green line represents the acceptable failure rate, while the red line represent the unacceptable failure rate. The black line represent the behaviour of a hospital.
In Figure 3 we clearly see why this analysis is almost useless with our data. It is true that the CUSUM chart assesses that the hospital “030DJQ” has a good behaviour, but we don’t need to run the chart on all the data. In every plot the black line goes out really fast from the boudaries, whereas CUSUM charts are expected to be re-started, and limits to be readjusted since the alert is out (when the black line goes above the red line).
As such, while CUSUM charts are very useful in hospitals monitoring, they are unfitting for the dataset we use.

4 Hospital performance in terms of mortality rate

As said before, between CUSUM charts and funnel plots, we find the latter to be really fitting our aim. In the following sections we will show the results of the analysis carried out on the dataset under study.

4.1 Hospital Standardized Mortality Ratio

Mortality rate is used to assess quality of care. However in order to carry out a fair comparison, features such as the severity of the events, distribution of comorbidities, age of patients must be taken in account. If proper adjustment are not performed unrealistic or biased conclusion may arise, see for example [27], where expert practitioners resulted to performe worse than newbies, because treating more difficult patients. For these reasons in our model we standardize the mortality and adjust for case mix of patients.

Indirect Standardization Method

A Hospital Standardized Mortality Ratio is calculated as the observed number of deaths divided by the expected number of deaths [27], given its case-mix of patients. The expected probabilities of in-hospital death are estimated using a logistic regression model with in-hospital mortality as outcome and patient features as covariates. All the data are usually used in this modelling. An expected probability of death is then evaluated for every patients in the dataset. This expected probability can be seen as the probability of in-hospital death for that patient in a standard hospital in Lombardia.
Given a hospital, the patients admitted in that hospital are selected. The sum of their expected probabilities gives the total expected number of deaths (E) for that hospital.
The observed number of deaths (O) is evaluated by simply counting the number of people who died in the hospital over the time period of interest. Finally the indirectly standardized mortality ratio is given by the ratio between the observed number of deaths and the expected number of deaths. For further details [27].

Logistic Models

Our dataset is event based and a patient can have more than one HF event. While events related to different patients can be considered as indipendent instances, the same assumption doesn't stand if looking at different HF events related to the same patient. For this reason the focus has to change: we divide the dataset using the hospitalization number as a label. We select all the data related to the first, second and third hospitalization for further studies. In the subset containing all the data related to a hospitalization k, there is a one to one relation between events and patients, because every patient can be admitted just once for its k^{th} time.
Our logistic models are fitted on the subsets of interest (one for each subset), with subsets beeing data related to first, second and third hospitalization events.
The models use as explanatory variables: Sex, Age, risk coefficient for comorbidity, Length of Stay, flags for Intensive Therapy, cardiochirurgy, Shock events, presence of other procedures, while the response variable is in-hospital death.
The result for the model fitted on first hospitalized patients is shown in Table 7.

	Model 1
(Intercept)	-7.18^{***} (0.10)
SexM	0.13^{***} (0.02)
Age	0.06^{***} (0.00)
Risk	0.20^{***} (0.00)
LOS	-0.02^{***} (0.00)
Intensive Therapy	0.74^{***} (0.02)
Cardiochirurgy	-0.43^{***} (0.07)
ICD	-1.89^{***} (0.19)
Shock Event	3.61^{***} (0.03)
PTCA	-0.58^{***} (0.04)
AIC	125271.02
BIC	125396.33
Log Likelihood	-62623.51
Deviance	125247.02
Num. obs.	253475

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Statistical models

Table 7: Summary for the logistic model fitted on patients at first re-hospitalization. The response variable is in-hospital death.

As we see in Table (7) all the coefficients are significant. That is expected, considering the number of patients examined. Coefficients related to LOS, presence of ICD, PTCA and Cardiochirurgy Care are negative, meaning that both a short length of stay and these procedures have a positive influence on the probability of survival. Instead the presence of an event of shock during the hospitalization strongly hinder the same probability. Risk, Intensive Therapy, beeing male and Age also increase probability of death. Age and the presence of a Shock event seem to be the strongest factors in decreasing probability of survival, while the ICD implant the strongest one in increasing it. The LOS coefficient is negative, but we notice that also rehabilitative admissions are considered, so this coefficient may be influenced by the rehabilitative patients.
We then model mortality rate on second and third hospitalized patients. In every model the c^2 statistic is good (≥ 0.75) and all the coefficients are significant.

4.2 Identification of hospital outliers via funnel Plot: mortality outcome

Mortality rate are standardized as explained in Section 4.1.
The funnel plots of Standardized Mortality Rate (SSR) for patients at their first hospitalization are shown in Figure 4. The funnel plot of SSR for patients at their second hospitalization is reported in Figure 6. While the funnel plot of SSR for patients at their third hospitalization is in Figure 7. In these Figures:

1. $Y = O/E$ is the measured indicator for every institution. O is the observed number of deaths and E the expected number of deaths.
2. The target SSR is $\theta_0 = 1$, since Observed deaths are assumed to be distributed as a *Poisson* with mean value E and variance θ/E.
3. The precision parameter ρ is given by $\rho = E$ and $g(\theta_0) = \theta_0$.
4. funnel limits are as in (2) or, if overdispersion correction is needed as in (5) or in (9).

In our funnel plots we identify as negative outliers the hospitals that are above the upper funnel limit (0.99). We identify as positive outliers the hospitals below the lower funnel limit (0.01). Finally a hospital is "under control" if it lies whithin the funnel limits.

4.3 Standardized mortality rate at first, second and third admission

Here funnel plots are applied to analyze hospital mortality rates.

Mortality at first hospitalization

Only first hospitalizations are considered: patients are 252,875, admitted in 232 hospitals. We consider now hospitals with at least 20 patients admitted for their first event. In Table 8 and 9 some summary statistics of data are reported.

	Number	
Patients	252,875	
Hospitals	232	
	Females	Males
Patients	53%	47%

	percentage
Dead	0.086
ReH	0.400
IT	0.20
Cardiology	0.053
ICD	0.012
DRG	0.15
Shock	0.024
PTCA	0.054
CABG	0.03
MDC5	0.99

Table 8: Summary of the dataset with hospitals with at least 20 first events.

	median	mean	SD
Age	79	77.15	11.05
Risk	1	0.99	1.42
LOS	10	13.67	14.19

Table 9: Summary of the dataset with hospitals with at least 20 “first events” each.

The funnel reported in Figure 4 shows the Standardized Mortality Ratio plotted against the Expected cases.

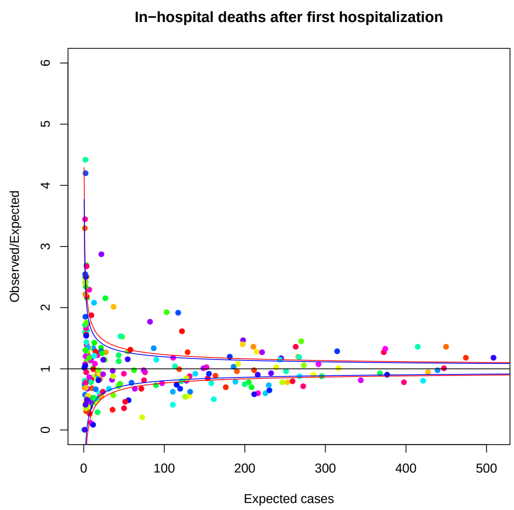


Figure 4: Funnel plot of Standardized Mortality Ratio evaluated on patients at first hospitalization. Funnel limits are computed without overdispersion correction, see (2).

We see that, as expected, dispersion around the funnel decreases as long as hospital dimension increases, but a large number of institutions lies outside the funnel. Since 39 hospitals (16.8%) are identified as “positive outliers”, 43 (18.5%) as negative outliers and only 150 as under control units we need to check for overdispersion.
As previously explained in Section 2.2, we assume overdispersion only if the estimate of ϕ is bigger than the threshold defined in (7). Since the threshold is equal to 1.19 and $\hat{\phi} = 6.1$, we assume overdispersion and we evaluate the funnel limits as specified in Section 2.2. We present them in Figure 5.

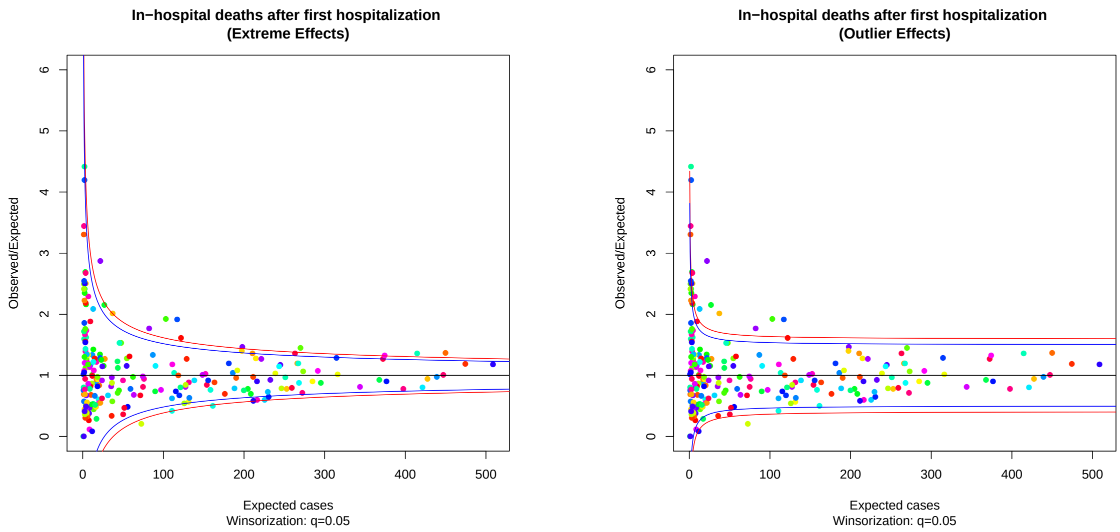


Figure 5: Funnel plot of Standardized Mortality Ratio evaluated on patients at first hospitalization. Funnel limits evaluated with multiplicative (left) and additive (right) correction, see (5) an (9) respectively.

Both funnel plots in Figure 5 show funnels with limits adjusted for overdispersion. The first funnel has limits evaluated with multiplicative correction (5), while the second funnel has limits evaluated with the additive random-effects correction as in (9). The funnel on the left in Figure 5 identify 2 structures (0.8%) as positive outliers and 11 (4.7%) as negative outliers, the right one identify 2 structures as positive outliers and 15 (6.4%) as negative outliers.
The two sets of identified outliers are a subset of the ones previously identified in Figure 4, but the two groups are very different from each other. Indeed it's easy to see that the two methods label outliers very differently: between the structures identified as outliers only 4 are labelled the same (3 negative and 1 positive). Funnel limits evaluated with extreme outlier effects, or multiplicative approach, are more lenient with smaller hospitals than with bigger ones. Most of the outliers identified with funnel limits evaluated with additive correction are indeed small structures.
If the aim of the monitoring is to identify only few health authorities as outliers, correction for overdispersion must be made. When adjustments for overdispersion are needed, they should be chosen accordingly to the aim of the monitoring. Indeed we can assess the subset of hospitals we need to focus our attention on: if small hospitals are to be monitored as strictly as bigger ones we can choose the additive overdispersion correction, if on the contratry we are more interested in outlier detection between bigger hospitals, the multiplicative approach is to be preferred.

Mortality at second hospitalization

Only second hospitalizations of patients are now considered. Hospitals included in the study have at least 20 patients that are admitted for their second event. In Tables 10 and 11 some summary statistics are reported.

	Number	
Patients	$100,204$	
Hospitals	187	
	Females	Males
Patients	50%	50%

	percentage
Dead	0.093
ReH	0.500
IT	0.150
Cardiology	0.035
ICD	0.032
DRG	0.140
Shock	0.022
PTCA	0.028
CABG	0.015
MDC5	0.830

Table 10: Summary of the dataset with hospitals with at least 20 "second events" each.

	median	mean	SD
Age	79	77.82	10.48
Risk	1	1.56	1.62
LOS	9	13.20	13.50

Table 11: Summary of the dataset with hospitals with at least 20 second events.

The funnel of Standardized Mortality Ratio evaluated on the dataset of patients at their second hospitalization shows overdispersion. While this overdispersion lessens, we have still 17 hospitals (9%) identified as "positive outliers", 19 (10.2%) as negative outliers and only 151 as under control units, so we check for overdispersion.
The estimate of ϕ is smaller than before, but bigger than the threshold defined in (7). Since the threshold is equal to 1.21 and $\hat{\phi} = 2.7$, we assume overdispersion and we evaluate the funnel limits as specified in Section 2.2. We present them in Figure 6.

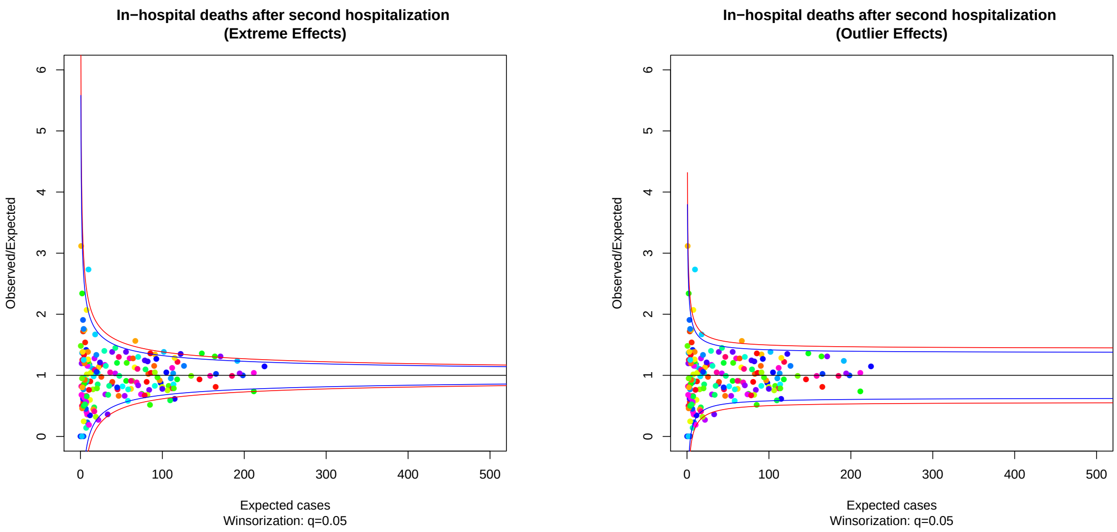


Figure 6: Funnel plot of Standardized Mortality Ratio evaluated on patients at second hospitalization. Funnel limits evaluated with multiplicative (left) and additive (right) correction, see (5) an (9) respectively.

As before the identified outliers remain a subset of the outliers identified with limits evaluated as in (2) and they remain very different from each other. Among the structures identified as outliers only 2 are labelled the same by the two methods. Funnel limits evaluated with extreme outlier effects, or multiplicative approach, are once again more lenient with smaller hospitals than with bigger ones. Most of the outliers identified with funnel limits evaluated with additive correction are instead small structures.

Mortality at third hospitalization: Only third hospitalizations of patients are now considered. Hospitals included in the study have at least 20 patients that are admitted for their third event. In Tables 12 and 13 some summary statistics are reported.

	Number	
Patients	50,046	
Hospitals	161	
	Females	Males
Patients	47%	53%

	mean
Dead	0.099
ReH	0.560
IT	0.140
Cardiology	0.026
ICD	0.037
DRG	0.130
Shock	0.022
PTCA	0.022
CABG	0.010
MDC5	0.830

Table 12: Summary of the dataset with hospitals with more than 20 “third events” each.

	median	mean	SD
Age	79	77.62	10.23
Risk	2	2	1.73
LOS	10	13.43	13.97

Table 13: Summary of the dataset with hospitals with more than 20 third events.

In Figure 7 we show the funnel plot of Standardized mortality evaluated on patients at their third hospitalization. To show clearly the decreasing of the overdispersion we report the plot where funnel limits are evaluated without correction. A brief comparison with Figure 4 points out that fewer hospitals are outside the funnel. Only 8 hospitals (5%) are identified as positive outliers and 11 (6.8%) as negative outlies. If we check for oversipersion we find that the estimate of ϕ is still bigger than the threshold defined in (7). The threshold is still around 1.2 and $\hat{\phi} = 1.9$. Since the two values are similar we show the funnel plot with limits without overdispersion correction.

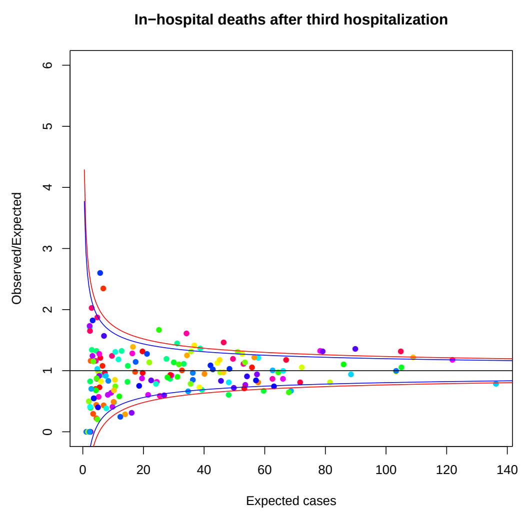


Figure 7: Funnel plot of Standardized Mortality Ratio evaluated on patients at third hospitalization. Funnel limits evaluated without overdispersion correction, see (2).

Remarks on outlier behaviour with hospitalization

Two remarks can be made about the outlier behaviour with increasing hospitalization. Firstly overdispersion decreases clearly from the first hospitalization to the third. Hospital Standardized Mortality Rate evaluated on patients at their first hospitalization are much more dispersed around the mean than ones evaluated on data related to second and third admissions. Since an overdispersed behaviour is usually due to the impact of unmeasured covariates that are not taken into account in any risk-adjustment method, the lessening of the overdispersion can be due to some kind of evening out of the patients at further hospitalizations. Secondly the outlier behaviour of an hospital doesn't change with different patient history: all the hospitals labelled as outliers for SSR evaluated on patients at their third admission, are also outlier at first hospitalization, while between first and second only one hospital exits the funnel limits.

Therefore we can say that while the number of outliers decreases, the ones that are outlier at third hospitalization are outliers also at first one. On the other hand hospitals which are within the funnel limits in patients at their first event, remain within the funnel limits also for further hospitalizations. So hospitals tend to pull back inside the funnel limits, while no hospital exits the funnel plot in further hospitalization if they weren't outside in the first one.

5 Hospital performance in terms of re-admission rate

While in Section 4 hospital performance is evaluated using mortality rate, a different perspective is now introduced. As previously stated, Hearth Failure is a chronic disease, where around 40% of patients are re-hospitalized after their first admission. Therefore, also outlier behaviour relating to the re-hospitalization rate can be monitored. As previously defined, the re-hospitalization rate is the rate of patients that will have another event after the current one. In this section funnel plots are constructed to analyze hospital re-hospitalization rates.
While the mortality rate was standardized, the re-hospitalization rate we analyze now isn't, but a standardization can be considered. The re-hospitalization rate is evaluated for each one of the three subsets used in Section 4.2.

5.1 Identification of hospital outliers via funnel Plot: re-hospitalization outcome

In Figures 8 and 9 are shown the funnel plot of re-hospitalization for patients at their first, second and third hospitalization. In these figures all funnel limits are evaluated without overdispersion correction.

1. $Y = p$ where p is the observed rate of rehospitalization.

2. The target θ_0 is the mean of the current dataset, that means that re-hospitalization rates are assumed to be distributed as a Bernoulli distribution with mean θ_0 and variance $\theta_0(1-\theta_0)/n$.

3. The precision parameter ρ is given by $\rho = n$ and $g(\theta_0) = \theta(1-\theta_0)$.

4. Funnel limits are as in (2) or, if overdispersion correction is needed as in (5) or in (9).

As before we identify as negative outliers the hospitals that are above the upper funnel limit (0.99), as positive outliers the hospitals below the lower funnel limit (0.01) and as "under control" hospitals that lie whithin the funnel limits.

5.2 Re - hospitalization rate at first, second and third admission

Funnel plot of re-hospitalization rate plotted against the number of events is reported in Figure 8. As with the mortality rate funnel plots, units identified as "out of control" are too many: 31 (13.4%) hospitals are labelled as positive outliers and 49 (21.1%) as negative outliers.

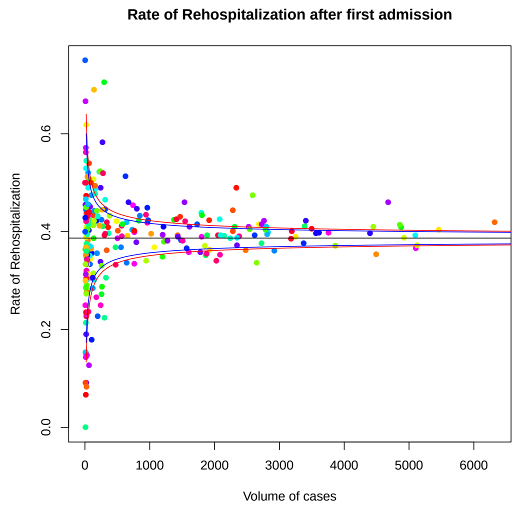


Figure 8: Funnel plot for re-hospitalization evaluated on patients at first hospitalization. Funnel limits evaluated without overdispersion correction, see (2).

As before overdispersion is present: indeed the estimate of ϕ is $\hat{\phi} = 4$, while the threshold defined in (7) is around 1.18. We then assume overdispersion and we evaluate the funnel limits as specified in Section 2.2. With the additive overdispersion correction 11 (4.5%) hospitals are identified as positive outliers and 12 (5.2%) as negative outliers. With the multiplicative correction all the hospitals but 4 (1.7%) negative outliers are considered under control. Only 3 hospitals are labelled as negative outliers by both methods. The great difference in labelling is present because most of the hospitals identified as negative outliers by funnel limits evaluated with the additive correction are once again small hospitals, that are within the funnel limits obtained with multiplicative correction.

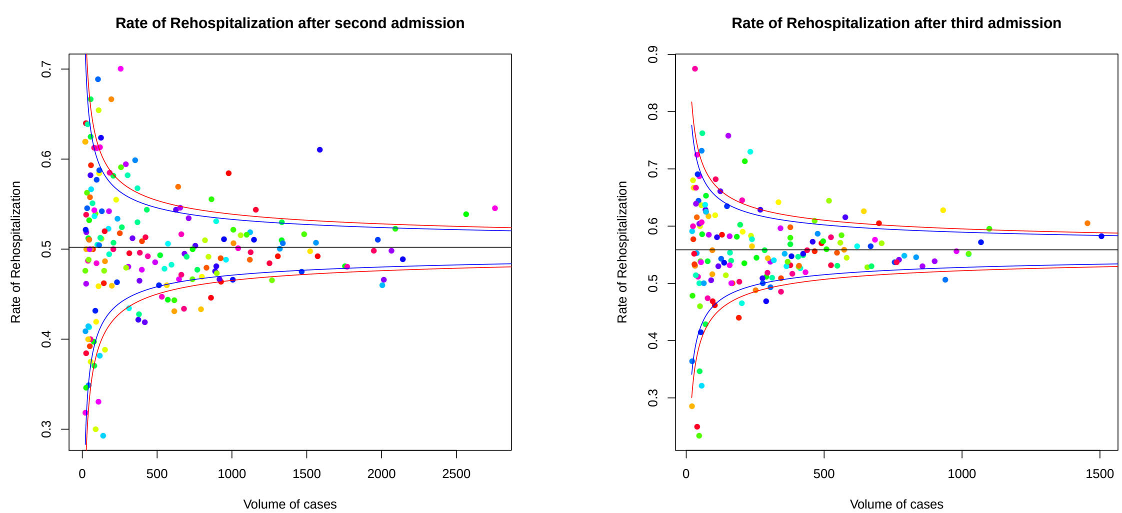


Figure 9: Funnel plot for re-hospitalization evaluated on patients at second and third hospitalization. Funnel limits evaluated without overdispersion correction, see (2).

In Figure 9 are shown the funnel plot of re-hospitalization rate for patients at their second and third hospitalization. Overdispersion ($\hat{\phi} > 2$) is still clear in both funnel plots relative and we correct the funnel limits as before. While we have evaluated funnel limits with overdispersion correction and we have identified the hospitals that are identified as outliers, in Figure 9 we show funnel limits drawn without correction so that the lessening of overdispesion can be clearly seen.
One (second admission) or no (third admission) outliers are obtained if the multiplicative correction for overdispersion is used. The funnel limits obtained with additive correction highlight the presence of 7 (3.7%) negative and 3 (1.6%) positive outliers at second admission, while 5 (3.1%) negative and 4 (2.5%) positive outliers are pointed out at third admission.
Two main remarks can be made about the outlier behaviour with hospitalization. Firstly overdispersion decreases from the first hospitalization to the third. Re-hopitalization rates evaluated on patients at their first hospitalization are more dispersed around the mean than ones evaluated on data related to second and third admissions. Secondly there is once again permanence of outliers between rehospitalization, as only one hospital identified as outlier in the analysis relative to the third hospitalization isn't between outliers highlighted at first hospitalization. Hospitals tend to pull back inside the funnel limits, but some of the outliers still remain outside as seen before.

Remarks

In Chapters 3,4,5 we presented funnel plots as a simple and effective way to monitor hospitals' performances. In fact, while beeing an interesting way to explain data on institutional comparisons, they're also an explicative one. With axis that are easily interpretable and the outliers as points outside the funnel limits, funnel plots are ideal to communicate data clearly. Furthermore as stated previously, there is no spurious ranking of institutions, additional variability in institutions with small volume is allowed and finally overdispersion can be taken into account.
While beeing an attractive communicative device, the funnel plot isn't appropriate to study the progress of individual institutions over time, since a basis for sequential testing due to repeated testing is not provided. Methods such as Cumulative Sum charts or risk-adjusted CUSUMs provide a formal basis for such sequential monitoring. In conclusion we suggest that CUSUM charts are to be used to monitor hospital on the run, while funnel plots can be used to evaluate the hospitals once in a while, likely once in a year.

Part III
Network analysis

In Part III we introduce the network analysis and explain why the hospitals can be considered as a group instead of different and separate units. We introduce the hospital networks we work with and the models we implement. For further reading on statistical analysis of network data we refer to the book from Kolaczyk and Csárdi [21]. Most of the analysis have been carried out using the “igraph” package of R [8] and the “ergm” package [14, 17]. In this part we reduce our attention to a subset of hospitals, choosing those (80 units) with at least 2000 patients.

6 Introduction to Graphs

A graph is a representation of a set of units where some of them are connected by links in pairs. The interconnected units are called vertices, and the links are called edges. A graph is graphically represented as a set of dots for the vertices, joined by lines that represent edges.
The edges may be directed or undirected. For example, if the vertices represent hospitals in our network in our hospital netwok, and there is an edge between two hospitals if they are similar, then this is an undirected graph, because if hospital A is similar to hospital B, then hospital B is also similar to hospital A. In contrast, if there is an edge from hospital A to hospital B when a patient from A is transferred to hospital B, then this graph is directed, because the transfer is not a symmetric relation. This latter type of graph is called a directed graph and the edges are called directed edges.

6.1 Definitions

A *graph* is an ordered pair $G = (V, E)$ where V denotes the set of vertices and E the set of edges. In our networks V and E are finite. The *order* of a graph is the number of vertices $|V|$ and the graph's *size* is the number of edges $|E|$.

Vertices: Vertices are the nodes of the graph, they can be connected in pairs, but they may also exist even without belonging to an edge. Given a directed edge that connect hospital i to hospital j, j is called the *head* (or hospital *out*) and i is called the *tail* (or hospital *in*).

Edges: Every element in E is an ordered (directed graph) or unordered (undirected graph) list of two vertices, these two vertices are the ones that the edge connects. The vertices belonging to an edge are also called the ends or end vertices of the edge. In a directed graph, given two vertices i and j, if there is an edge from i to j and an edge between j and i, the two edges are defined *mutual*. An edge can be also called *tie* or *link*.
A graph is said to be *weighted* if a number (weight) is assigned to each edge. Such weights can represent for example flows between two units, costs or distances.

Connection, transitivity and mutuality: In an undirected graph G, two vertices i and j are *reachable* if there is a path from i to j.
A graph is called *connected* if every vertex is reachable from every other. A directed graph is called *weakly connected* if replacing all of its directed edges with undirected edges produces a connected graph. It is *strongly connected* if it contains a directed path from i to j and a directed path from j to i for every pair of vertices i, j.
The *transitivity* of a graph is based on the relative number of triangles in the graph, compared to total number of connected triples of nodes. Where a triangle consists of 3 nodes that are completely connected to each other and a connected triple consists of three nodes i, j, k such

that node i is connected to node j and node j is connected to node k. The factor of 3 arises because each triangle gets counted 3 times in a connected triple. It's computed as:

$$T = 3 \cdot \frac{t}{CT} \tag{10}$$

Where t is the number of triangles in the observed network and CT is the number of connected triples of nodes in the network. $0 \leq T \leq 1$.
Mutuality in a directed graph is the percentage of edges that are mutual.

Degree and Strength: The *degree* of a vertex is the number of incident edges. In an undirected graph there is no distinction between out-degree and in degree, while in a directed graph the *out-degree* in the number of edges that exits from the vertex and the *in-degree* is the number of edges that enters the vertex.
The *strength* of a vertex is the sum of the weights of incident edges. *Out-strenght* and *In-strenght* correspons to Out-degree and In-degree.

Centrality: *Centrality* is a vertex features. There are various measure of centrality, and are all linked with the strength and the importance of an hospital in the network. Examining the centrality of a structure is like asking how critical that structure is in the general flow or connection. Measures of centrality are designed to quantify various notion of "importance" of a node.
There are a vast number of centrality measures that have been proposed (see [21] for details). The most widely used remain the vertex degree, that we have already described above.
Betweennes centrality summarizes the extent to which a vertex is located between other pairs of vertices. This centrality measure highlights the importance that relates to where a vertex is located with respect to the paths in the network. The most commonly used betweeness centrality measure, introduced by Freeman [11], is defined as:

$$C_B(v) = \sum_{s \neq t \neq v \epsilon V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \tag{11}$$

where $\sigma(s,t|v)$ is the total number of shortest paths between s and t that pass throught v and $\sigma(s,t)$ is the total number of shortest paths between s and t. If the shortest paths are unique, the Betweennes centrality measures the number of shortest paths that pass throught v. In the following (Section 7) we evaluate the centrality of the nodes belonging to the network, as well as vertices' strenght and degree.

Layout:

The depiction of a graph must balance different aspects. The position of vertices and the length of edges must be chosen so that the graphs avoids confusion at its best. To watch out for graphic aspects it's necessary: nodes must be placed so that all the edges are of more or less equal length and there are as few crossing edges as possible. This is done by assigning forces among the set of edges and the set of nodes, based on their relative positions, and then using these forces to minimize their energy.
There are several algorithm that one can use to generate this equilibria: we mostly use the Kamada Kaway layout [20]. This approach is automatically implemented in the "igraph" package [8] automatically. For reasons of random initialization, every time the same network is plotted the result differs slightly (i.e: the postioning is different).

6.2 Network as the right perspective

Network analysis has been used for years, mainly to advance research in the social sciences [2], but has been developing faster in recent years ([8], [9], [12], [6], [17], [24]) due to the recent advances in technologies. In fact, internet and and especially on-line social networks produce a great deal of relational data, i.e., data which need to be analyzed considering their intertwining nature. The network perspective enables to model the relationships among units, considering patterns and pattern implications. As network approach is the most natural way to analyze this kind of interactive data, the research in the area is lively.
As people in a social network are connected by data flow, institutions can be considered as units connected by human flow. Networks are an unique tool to extrapolate flows information and information about unit authority or centrality. Understanding the patients flow in an hospital network is surely of interest, since it can highlight an additional outlier behavior, such as excessive out-flow or in-flow of patients. A flow network is the first hospital network we explore. Other hospital networks can be proposed: as an example in Section 7.2 we introduce a hospital network where similarity between hospitals is accounted for on edges.
Two are the main issues when dealing with modelling a network. First of all, one must choose properly the variables to be represented on nodes and edges, as well as the weights the edges must be labelled with. This is a descriptive step of the analysis, that allows, if properly approached, for useful insights of the problem under study. A second issue is how to fit a model on the chosen graph.
In Section 7 we describe the hospital network we propose. In Section 8 we present the generalized linear models we can fit on them.

7 Networks of Hospitals

In this section we propose how to look at hospital networks within Lombardia in two different ways. Firstly we state our choices regarding vertices, edges and their representation, then we analyze the proposed networks.

7.1 Vertices

The first step is to label the vertices. Since in every proposed network vertices are alike, we present them beforehand.
As we previously stated, every hospital is not completely indipendent of others. Several connections between structures can be made and are proposed in what follows, however all the networks we study see hospitals as vertices. Hospital features are inserted as nodes attributes, according to instructions in package “igraph” of R [8]. Some features are displayed in terms of nodes colors, size, shape, among others. For each institution, in addiction to the identification code, the features reported in Table 14 are available.

(Standardized Mortality rate Color)	(Re-hospitalization rate Color)
Standadized Mortality Ratio (SSR)	Re-hospitalization Rate (reH)
mean risk	Number of patients (N)
median Lenght Of Stay (LOS)	mean expense (worth)
Cardiochirurgy percentage	ICD percentage
Intensive Therapy percentage	PTCA percentage
DRG percentage	Shock event percentage
CABG percentage	MDC51 percentage

Table 14: Vertex features in the hospital network. Indicators measured in Funnel Plots described in Section 4 and 5 are used to color the nodes.

As we said informations relative to the vertices can be displayed in the graph. As such the outlier labels obtained with funnel plots are recorded in "Mortality Rate Color" and "Re-hospitalization rate Color" and used to color the network. "Positive outliers" (Green), "under control units" (Yellow) and "negative outliers" (Red) can then be easily identified in a graphical way. The color can vary between the color obtained by the mortality funnel plot (in Table 14: "standardized Mortality Rate Color") and by the rehospitalization one (in Table 14: "Re-hospitalization rate Color") at need.
In particular we considered funnel plots with limits evaluated without overdispersion correction to determine the color of vertices. Other choices can be made, using the outliers found by other limits, depending on the aim, as stated before.
We choose median Length Of Stay instead of the mean to detect the rehabilitative admissions effect. Since we are interested in the hospitals dimission policy, we prefer to deal with typical admissions and not rehabilitative ones. Lastly the "mean worth" is the mean expense of admissions in the hospital.
As two other features can be shown in figures as every vertex's width and height, we choose vertex dimensions proportional to hospital dimension (Width) and mean worth (Height).

7.2 Hospital network as a flow graph

The first hospital network see the flow of patients on the edges. Patients that have more than one hospitalization either go every time in the same hospital, or they can change hospital. If they change hospital we have a patient that moves along an edge.
To create this kind of graph we join each hospital with all the others with a "potential" directed edge. The resulting graph is a complete directed graph. For every patient we check his/her admissions, if the patient has an hospitalization in hospital A and the following in the hospital B, we add plus one to the edge between A and B. Null edges between hospitals are then deleted.

Edges

We must highlight the fact that this kind of patient flow is really small in numbers. On the first hand we know that more than half of the patients had just one admission and on the other hand the majority of the others tend to stick to the same hospital. Patients remain mainly in the same hospital, probably both for geographical reasons (the majority of chronic patients are elderly, so its really unlikely they go to a far away hospital) and because people prefer to stick to the same doctor or team if possible. So these edges completely ignore more than two third of the patients. These edges measure more or less the way chronic patients move between hospitalization and are strongly related with the geographical position, a variable that we are not allowed to measure.

Nontheless with this approach we can identify hospitals that are in-stars and out-stars for patients and also point out flows that clinician can identify as uncommon. An analysis on such a graph can give useful insights on the epidemiology of transfers and help with logistic issues that can arise.

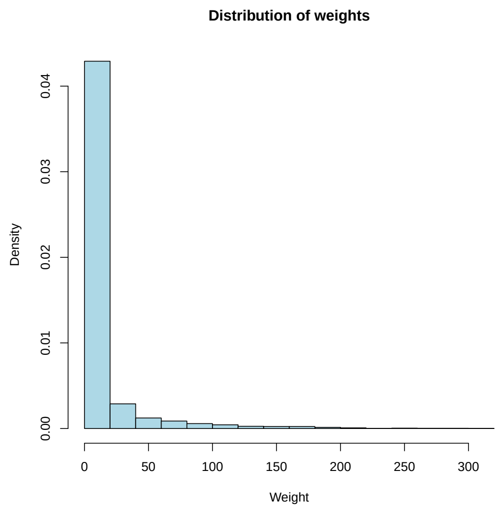


Figure 10: Weight distribution in the flow graph

As we can see in Figure 10 we have the great majority of really small edges (76% weight between 1 and 10). Since a nearly complete graph is difficult to represent, in the descriptive analysis that follows we use only edges that had more than 10 patients.

Graph representation

In Figure 11 we colour the hospitals by the Standardized Mortality rate color and in Figure 12) with the re-hospitalization one (see table 14).

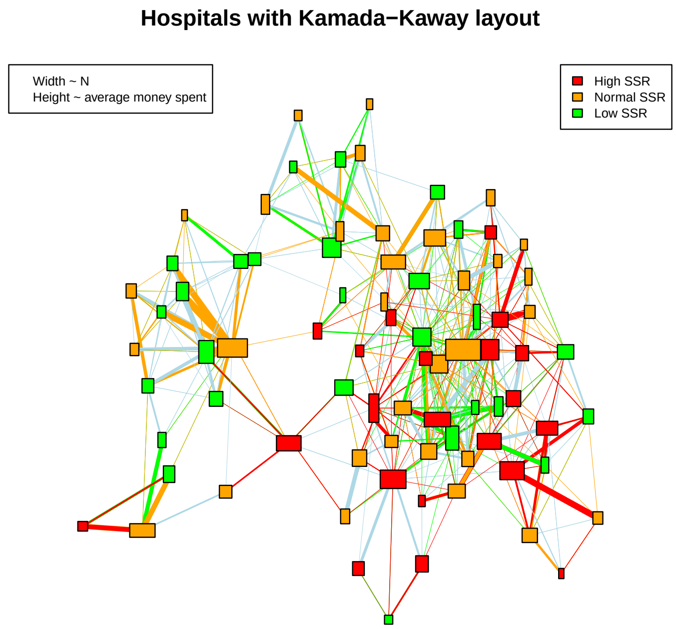


Figure 11: Hospital Network as a flow graph. Vertices are colored according to the SSR of the corresponding hospital, classified as “positive” (green) if under the lower limit of the Funnel Plot in Figure 4, “under control” (orange) if within the funnel limits in Figure 4 and “negative” (red) if above the upper funnel limit of the Funnel Plot in Figure 4. Edges are depicted as light blue if the flow is betwen same-color hospitals or colored as the “head” if between different-color hospitals.

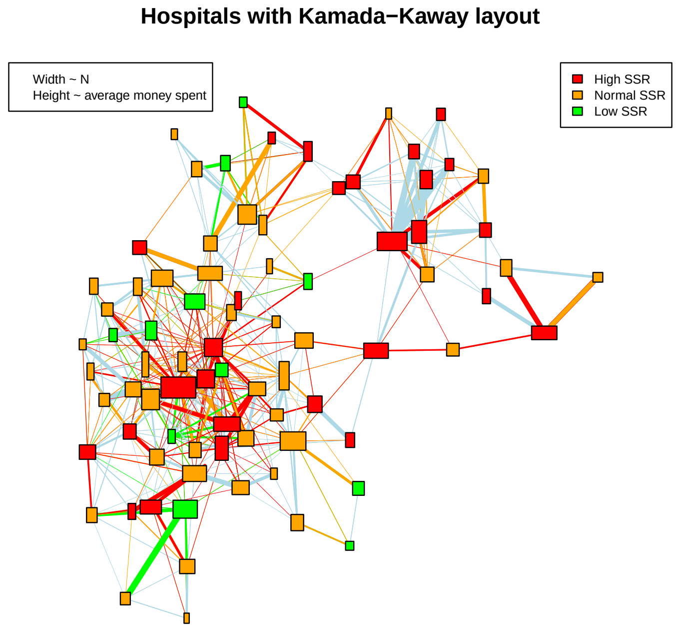


Figure 12: Hospital Network as a flow graph. Vertices are colored according to the SSR of the corresponding hospital, classified as “positive” (green) if under the lower limit of the Funnel Plot in Figure 8, “under control” (orange) if within the funnel limits in Figure 8 and “negative” (red) if above the upper funnel limit of the Funnel Plot in Figure 8. Edges are depicted as light blue if the flow is betwen same-color hospitals or colored as the “head” if between different-color hospitals.

The network we have previoulsy explained are reported in Figure 11 and Figure 12. The choice of features we display on vertices (number of patients, average money spent and outlyingness) can be changed at need. The current choice enables an efficient monitoring of the most important hospital indicators related to hospital efficiency. In both figures small structures with very high mean expence are easily spotted. When small numbers, high mean espense and underperformance behaviour coexist in a hospital, further analysis are needed. While this seem to be only a graphical aid, it can be worthwile when explaining data.
As a first step we examine the network to see what observations can be made about patients flow.
We are interest in the edge disposition, we want to know if smaller structures are clustered aroud bigger ones, if the flow is mainly between hospital of the same dimension and so on. If there are some kind of grouping and if we can identify differences between groups.
In Figures 11 and 12 we notice that:

- Most of the bigger edges are between hospitals that differs greatly in exposure.
- There some kind of clustering. It’s possible due to geographical reasons, but we are not allowed to verify.
- The mortality outlier distribution in Figure 11 seem to vary in different areas. So there

seem to be a clustering of the labels. The only group induced by color is on the left, where a group of green hospitals stays isolated from the rest. Instead we don't see a major effect of the color on to the edges dimension.

- The rehospitalization outlier distribution in the graph seem to vary in different areas, with red outliers condensed firstly in the top right, and secondly in the center of the bottom left group. Otherwise from Figure 11 hospitals have the majority of their neighbours of their color and outlier seem to be distributed as annular rings around a center of red outliers.
- The difference in coloring is easily spotted after noticing that the two graphs mirror on each other. The coloring done with mortality and with re-hospitalization are really different, almost contrary. This phenomena is really relevant in the group of hospitals that seems to be separated from the others, having an extremely good behaviour with respect to mortality as an extremely bad one with respect to re-hospitalization.

Vertices and edges features - Mortality

In what follows, vertices and edges features related to the network graph colored by Standardized Mortality rate are further analyzed.

In and out strength: While the degree of a hospital is the number of hospitals connected to it, the strenght is the number of patients that depart or arrive in the hospital. Both are indices of centrality and are linked with the hospitals importance. Unfortunately also geographical influence is concerned, since patient mobility among hospitals in the same city differs greatly from patients mobility between cities. Patients in a big city can easily change hospital staying inside the city itself, while in a small city there can be only one hospital. Since for elderly patients changing city is challenging, a latent geographical effect probably influences the strenght.
As the number of patients that flow between the hospital is the focus, we report streght plots.
In Figure 13 we see the relation between flow and dimension. The dependence of Out-strenght on the number of patients in a hospital is to be expected, since a great number of patients mean more patients that can move from the hospital itself. Instead the dependence of In-strenght on the number of patients is more difficult to explain, since the number of patients seem to be related the attractivness of the hospital.
Despite the clear dependence between dimension and flow neither among the two figures suggests that being an outlier modify the flow in any way. Hospitals behave in the same way even though they are labelled differently. In the end there doesn't seem to be a linear relation between flow and SSR.

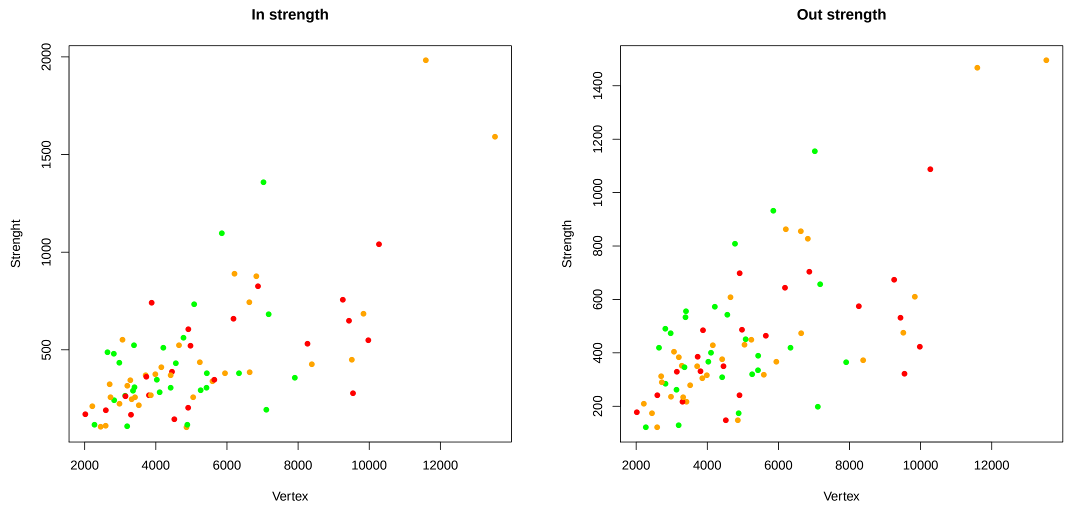


Figure 13: Plot of In-Strenght (left) and of Out-Strenght (right) for all the vertices in the graph flow. Vertices are colored accordingly to the corresponding SSR labelling.

Relations between outlier labeling and edges: In Figure 11 we notice that hospitals are neighbour with hospitals of every color, so once again there doesn't seem to be a relation between flow and SSR labelling. To identify if vertices have indeed some kind of preference we evaluate a relative same-color strength of vertices: the percentage of strength that is related to vertices of the same color. Given a vertex i of color K, with $K = \{Red, Green, Yellow\}$:

$$RSCS_K(i) = \frac{S_K(i)}{S(i)} \tag{12}$$

Where $S_K(i)$ is the strength (In or out) evaluated only considering hospitals colored "K" and $S(i)$ is the strength introduced before. In Tables 15 and 16 are reported the relative same-color strength of vertices.

Same Color Out-Strength					
Color Vertices	1st Qu.	Median	Mean	3rd Qu.	Max.
Red	0.140	0.310	0.330	0.470	0.890
Green	0.180	0.280	0.290	0.410	0.550
Yellow	0.260	0.350	0.410	0.590	0.820

Table 15: Summary table for same color Out-Strength for vertices in the Flow graph. Vertices are colored accordingly to the corresponding SSR.

Same Color In-Strength					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Red	0.140	0.320	0.320	0.490	0.770
Green	0.210	0.280	0.310	0.450	0.620
Yellow	0.240	0.380	0.400	0.530	0.860

Table 16: Summary table for same color In-Strength for vertices in the Flow graph. Vertices are colored accordingly to the corresponding SSR.

We see that the vertices strength are related to the percentage of colors in the graph (with 34% red, 40%, yellow, 26% green vertices). These results doesn't support a relation between SSR and flow. A weak relation can always be masked easily by geographical influence, that we aren't able to see.

Connection, transitivity and mutuality: In our data a lot of edges are mutual and the transitivity is high. In fact 89% of the edges are indeed mutual and the transitivity is also high. Moreover our graph is not only weakly, but also strongly connected according to the definition given in Section 6.1.

Vertices and edges - Rehospitalization

Here vertex and edges features related to the network graph colored by Re-hospitalization Rate are further analyzed.

In and out strength: In Figure 14 we see the relation between flow and dimension. The dependence of Out-strenght and In-strenght to the number is the same as for the mortality coloring.
Once again despite the clear dependence between dimension and flow neither of the two figures suggests that being an outlier modify the flow in any way. Hospitals behave in the same way even though they are labelled differently. In the end we see that beeing an outlier with respect to rehospitalization seems to increase Out-strength.

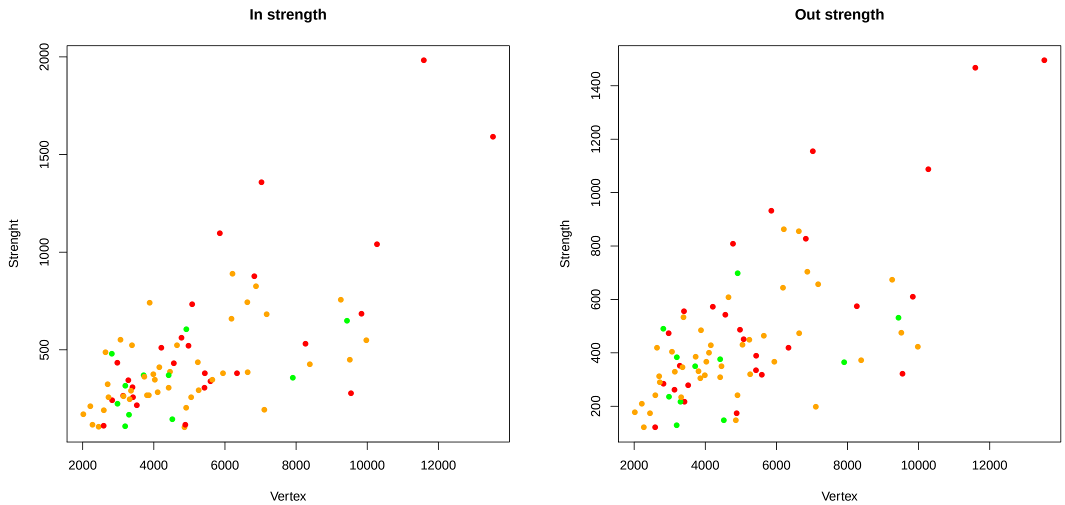
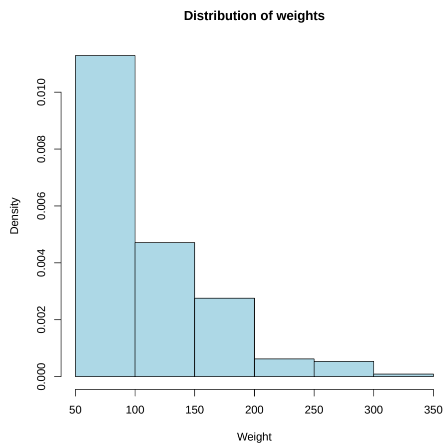


Figure 14: Plot of In-Strenght (left) and of Out-Strenght (right) for all the vertices in the graph flow. Vertices are colored accordingly to the corresponding Re-hospitalization Rate labelling.

Small edges as a noise

We stated before that very small edges are the majority, but with more than 2000 patients in 12 years, some patients can change hospital without any real pattern. In this section we reduce our attention to main patterns. We now build a graph that mantains only edges that weight at least 50 and we examine the network with colored by Standardized Mortality Rate.
It is likely that a latent geographical clustering is shown in Figure 15, but we are not aware of it (and it cannot be recovered) due to the encrypted identification key for hospitals. Nevertheless, it can be observed that some clusters behave better than others in terms of SSR.



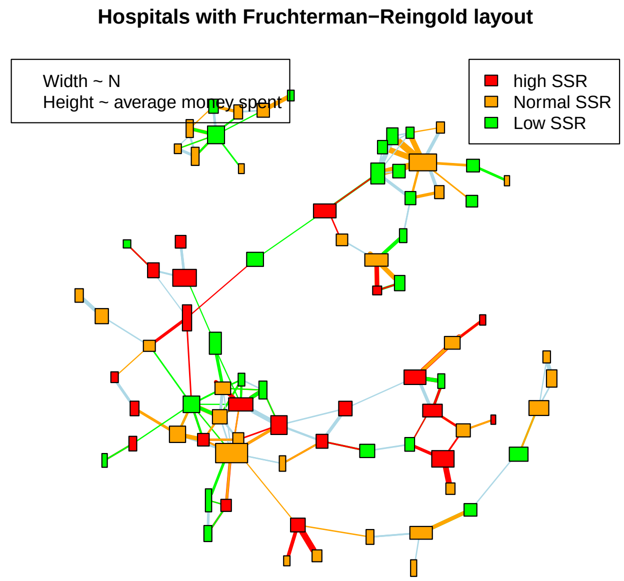


Figure 15: Distribution of weights for edges in the Flow Graph obtained cutting edges smaller than 50 on the left. Flow Graph obtained cutting edges smaller than 50 on the right. In the graph vertices color is chosen according to the labelling relative to SSR, while edges are depicted as light blue if the flow is betwen same-color hospitals or colored as the "head" if between different-color hospitals.

In Figure 15 we easily see that the majority of the edges are between hospitals with different labels. In Table 17 we show the vertices connected by the biggest edges.

	hosp_{in}	hosp_{out}	weight	color_{in}	N_{in}	color_{out}	N_{out}
3334	030ENR	030DKT	308	green	$4,780$	orange	$11,596$
6099	030JLQ	030DKT	285	green	$5,859$	orange	$11,596$
4284	030IJN	030DKV	283	orange	$4,157$	red	$6,186$
1397	030DKV	030IJN	275	red	$6,186$	orange	$4,157$
5860	030JFV	030DKR	261	orange	$3,861$	red	$9,435$
1262	030DKT	030JLQ	260	orange	$11,596$	green	$5,859$

Table 17: List of the vertices connected by the biggest edges in the flow graph of Figure 15 with weight and coloring of the structures al both ends of the edge. The coloration is done with the information derived from the funnel plot related to Mortality.

Unlike for the complete graph, SSR differences seem to be related to the flow. When we propose a generalized linear model to model the patients flow we will consider the difference in the variable regarding the mortality as a possible esplanation for the dimension of the edge between two hospitals.

Grouping in flow graphs

A question of interest is whether a graph separates into distinct subgraphs and, if this is the case, which are determinants driving such clustering. In our problem it's possible that the grouping is linked with the geographical clustering. Despite this previous knowledge, it can be of interest to study if there are some hospitals connected to others or if there are big flows between distinct groups. The direction of these flows is also of interest.

In the analysis of network graphs, clustering equals to find subsets of vertices that seems better "cohesive" in themselves than in the overall network. A "cohesive" subset is a group of vertices that are well connected among themselves and at the same time relatively separated from the other vertices. This problem is also referred as community detection. Many methods for graph partitioning are available and are well discussed in [24, 23, 7] to which we refer for further reading. We use primarily the "infomap community algorithm", a community detection algorithm for directed and weighted graphs [29].

We show in Figure 16 and 17 the results of grouping obtained with this algorithm, in such figures all the edges are of the same thickness and the coloring is once again the SSR one.

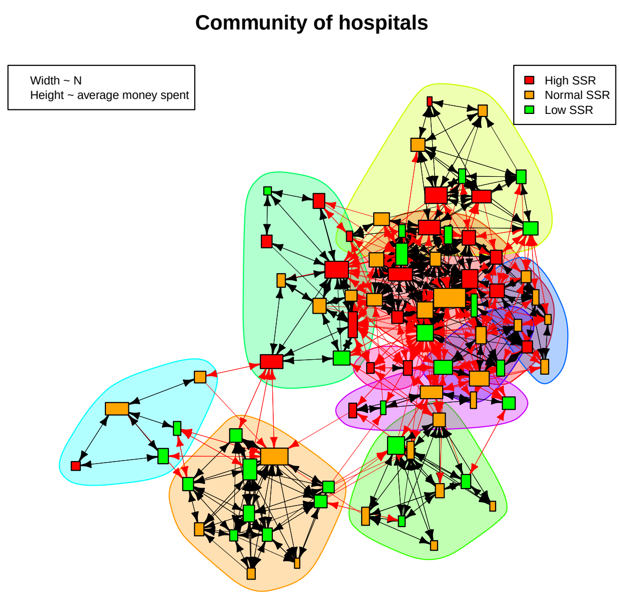


Figure 16: Flow Graph obtained keeping only edges bigger than 10. Vertices color is chosen with the labelling relative to Standardized Mortality Rate. Edges are all of the same thickness. edges color is black for intra-group edges, red for the others.

In Figures 16 and 17 black edges are intra group edges, while red edges are edges between different groups. Vertices color is chosen with the labelling relative to SSR, but the same analysis can be done with the labelling relative to R-hospitalization Rate. This choice let us see that the colors for mortality seem to be related to the clustering. As the number of groups is high, we won't test community for differences, but we will use the group information in the modelling that follows in Section 8.

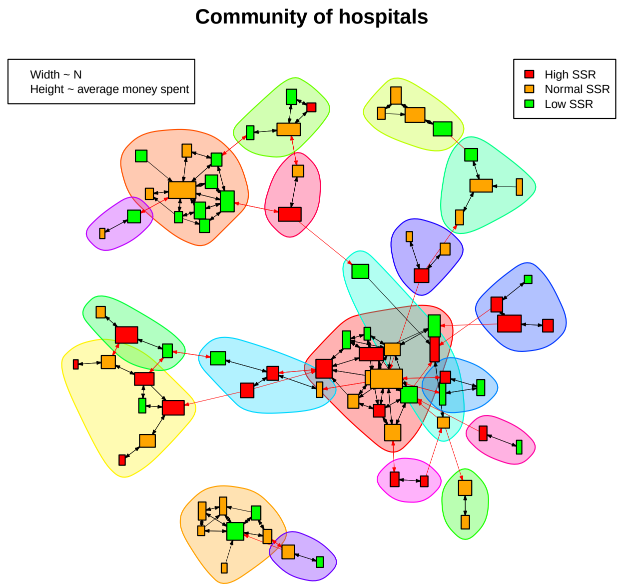


Figure 17: Flow Graph obtained keeping only edges bigger than 50. Vertices color is chosen with the labelling relative to Standardized Mortality Rate. Edges are all of the same thickness. edges color is black for intra-group edges, red for the others.

7.3 Hospital network as a similarity graph

In this section we propose a different way to look at the hospital network. The aim is now to look at a network where similar units are linked one each other, in order to explore if similar structures perform in a similar way. Now edges represent some kind of "distance"/ "similarity" between hospitals.
Clinicians drove the choice of features to be used to construct a suitable similarity measure. At the end we selected the following ones: the presence or absence of three medical facilities such as Cardiochirurgy, Hemodynamic and Intensive Therapy, and three continuous variables: global number of treated patients, global mean worth and median LOS.
The "Cardiochirurgy", "Emodinamic", "Intensive Therapy" are flags that mark the presence or absence of the related medical unit. The flags are extracted from the data: in assessing the presence of the Cardiochirurgy in a hospital we check that the sum of flags "cardio" relative to that hospital's patients is different from zero. For the Emodinamic we use the flags relative to "PTCA" and for Intensive Theraphy units we use the "IT" flags (see Table 2). Concerning the continuous variables, the total number of patients the hospital admitted is a measure of the exposure, the median LOS of patients is an index of the LOS policy of the structure, and the global mean worth for patient is an index of the economic policy. Global number of patients and Global mean worth are evaluated on the Complete Dataset.
For every pair of hospital x and y their distance is defined as follows:

$$d(x, y) = \sum_{i=1}^{6} \frac{x_i - y_i}{max_{(x,y)}(x_i - y_i)} \tag{13}$$

where:

$$x = (Globa\, mean\, worth, median\, los, Global\ N, Cardiochirurgy, Hemodynamics, IT) \tag{14}$$

Every hospital is now connected with all the others with a number that indicate the proximity. Small values are then to be interpreted as indices of similarity, whereas high values as indexes of dissimilarities, while high values are an index of dissimilarity. In Figure 18 the edge weights distribution is shown.

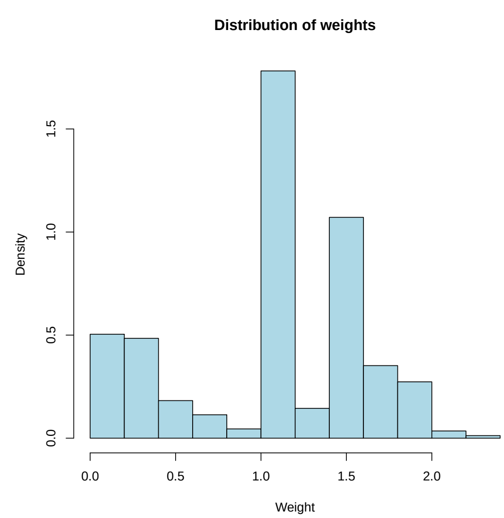


Figure 18: Weight distribution in the similarity graph obtained with distance 13.

As we see in Figure 18 there are three kinds of edges. Edges with weights between zero and one are a first group, edges with weights between 1 and 1.4 are in a second group and the third edges with weights greater than 1.4 are in the third.
Two hospitals may have a distance smaller than one only if they have the same flags. Two hospitals have a distance included between 1 and 1.4 either if they have really close similarity for all parameters except one of the flags or if they have the same structures but are rather different in at least two continuous variables.
Two hospitals with distance bigger than 1.4 might either have the same medical structures, but beeing very different for some others characteristics or could have different medical structures and not much similarity for other features, or both.
We now propose two possible threshold for edges, where hospitals bigger than the threshold are considered completely different.

Clustering of medical structures

In the first case we choose the threshold equal one so that we arbitrarily cluster hospitals for medical structures. What we actually do is artificially delete the edges whose values are bigger than one. If we mantain only edges with weight less than one, we group hospitals for medical structures: in the same group we have only hospitals that have the same equipment in terms of Cardiochirurgy, Intensive Therapy and Hemodynamics units and have enough similarity in the other three characteristics. Nearly all of the structure in the same group have the same degree. Our main interest is in assessing whether the natural clustering for medical structure is somewhat connected with the performance in terms of SSR or Re-hospitalization Rate. In Figure 19 we color the vertices using results from the funnel plot about mortality.

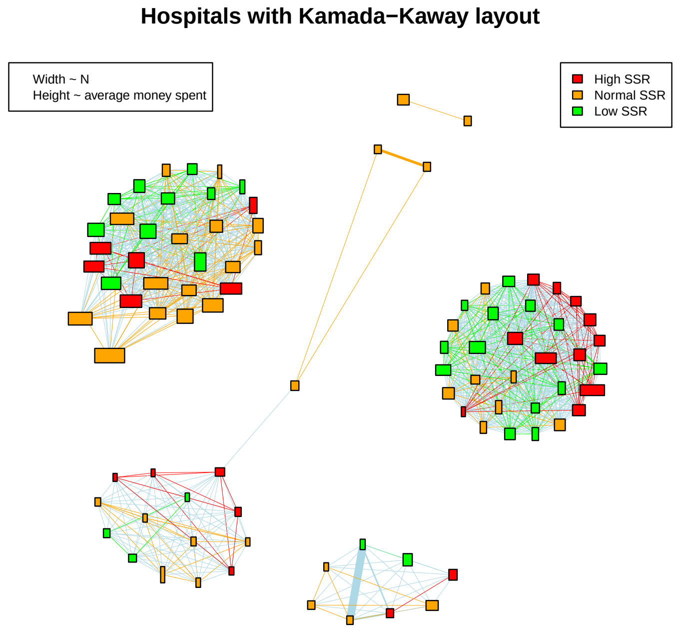


Figure 19: Graph obtained with weight computed by (13). We keep only edges smaller than 1. Vertices color is chosen with the labelling relative to Standardized Mortality Rate. Edge thickness is proportional to the similarity.

In Figure 19 we identify the group on the top left as hospitals that have all the medical units, while the group on the left is made by hospitals with both Intensive Theraphy and Emodinamic, but without Cardiochirurgy. All following comparison will be made between these two groups. In Figure 19 we also see that hospitals within the group with all the medical facilities are bigger that the ones in the other group. While the obtained clustering depends mainly on the flags, the two groups resulted different using the Wilcoxon test for all the variables used in the definition of (13).

In Figures 19 we notice also that the clustering by structures seem to be entirely independent of the "color" distribution. Testing the two bigger groups for difference in mean risk, SSR and re-hospitalization percentage let us to the same conclusion. So presence of a medical structure doesn't seem significantly linked with an improvement in the performance.

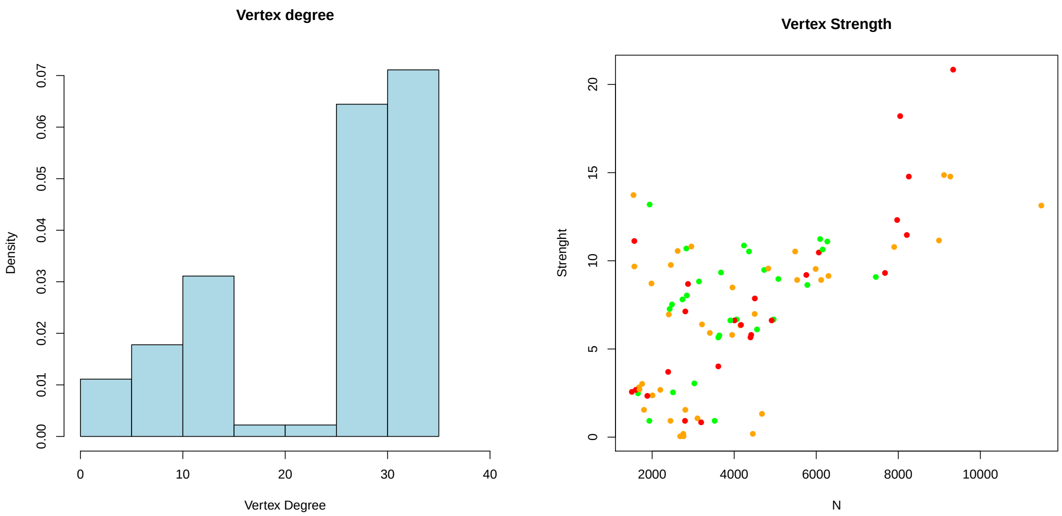


Figure 20: Histogram of Vertex degree for all the vertices in the graph in Figure 19 on the left. Plot of Strenght for all the vertices in the graph on the right. The coloring is done with the information derived from the funnel plot related to SSR.

Vertex characteristics, such as degree and strength (see Figure 20), highlight what we see in the networks in Figure 19, where hospitals in the same group have about the same degree. Furthermore there seem to be a connection between the strength and the dimension of the hospitals that is shown in 20, but this is probably due to some of the smaller hospitals having less connections (the smaller groups in Figure 19 are all made by small hospitals).

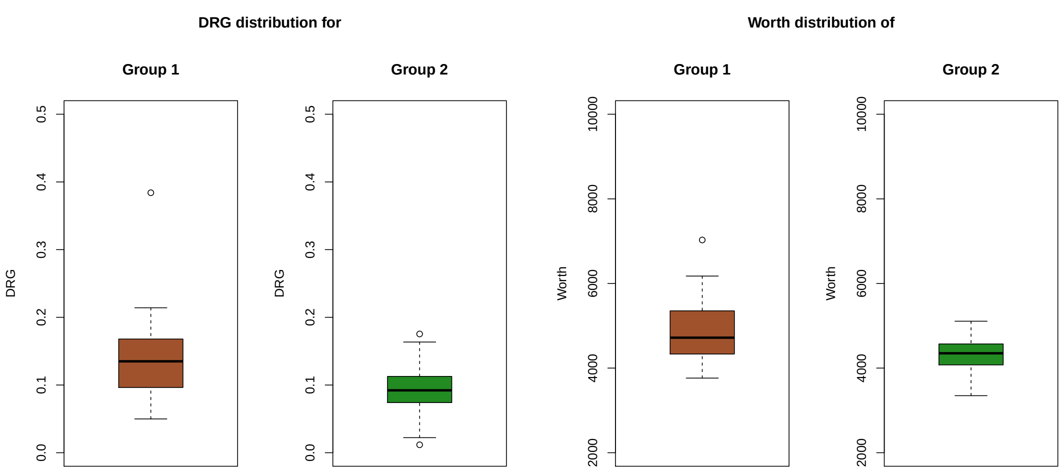


Figure 21: Boxplot for DRG (left) and mean expense (right) distribution in the two major groups. In each boxplot on the left the group with all the medical facilities, on the right the group without Intensive Teraphy Care. The corresponding graph is the one in Figure 26.

While the two major groups test the same as for the SSR and the rehospitalization percentage, they differed for the DRG ($p-value = 0.001$ and $p-value = 0.002$ with Willcoxon test). This difference is probably due to the DRG and worth distribution beeing almost the same (see 26). The DRG, beeing a flag that specify if the patient is evaluated as chirurgical or not, is indeed meaningful in the refound policy.

Different cutoffs

The cutoff choice is a crucial point in defining the possible features to be extrapolated from the graph. Whereas our previous choice (distance cutoff equal to 1) has a straightforward meaning for the analysis, other possible choices are not as easy to be interpested as it was. Looking at Figure 18 we see that another natural cut off is around 1.4 and in Figure 22 we see the graph that result if we put to zero all the edges bigger than 1.4. It's easy to see that this graph is completely different from the previous one.

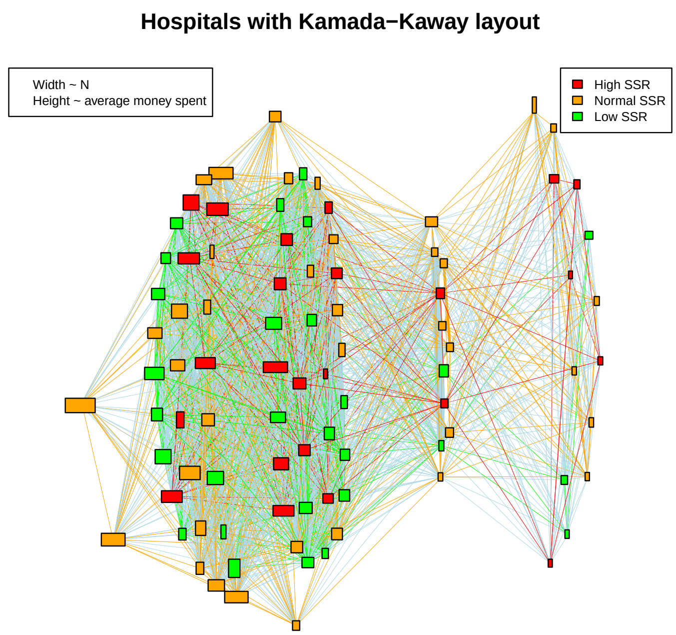


Figure 22: Graph obtained keeping only edges smaller than 1.4. The coloration is done with the information derived from the funnel plot related to Mortality.

The network in Figure 22 is conneced, but some kind of clustering is shown nontheless. As such we search for communities and five groups are identifyed by the fastgreedy.community algorithm [7]. While the flags still hold a great importance in assessing the membership, hospitals with a difference in flags are now allowed to be in the same group, but only if other similarity are really strong. The two main groups we identifty are still the ones we found previously, so for remarks and analysis we refer to Section 7.3.

Countinuous distance

As shown in Section 7.3 the distance previously evaluated (see 13) is strongly polarized by the three flags chosen. We decided to modify this index of dissimilarity and we asked clinician what continuous variables could be chosen to assess similarity. They suggested the use of $PTCA$, $cardiochirurgy$ and $Intensive\,Teraphy$ percentage to evaluated the distance between two hospital instead of the flags, as they represent a kind of procedural similarity. So for for every pair of hospital x and y their distance is evaluate as follows:

$$d(x,y) = \sum_{1}^{6} \frac{x_i - y_i}{max_{(x,y)}(x_i - y_i)} \tag{15}$$

where:

$$x = (mean\,worth, median\,los, N, p\,cardio, p\,ptca, p\,IT) \tag{16}$$

We evaluate all the distances between hospitals and we have yet again every hospital connected with all the others with a number that indicate the proximity.
In Figure 23 is shown the edge weights distribution and the resulting graph where edges bigger than the mean are deleted.

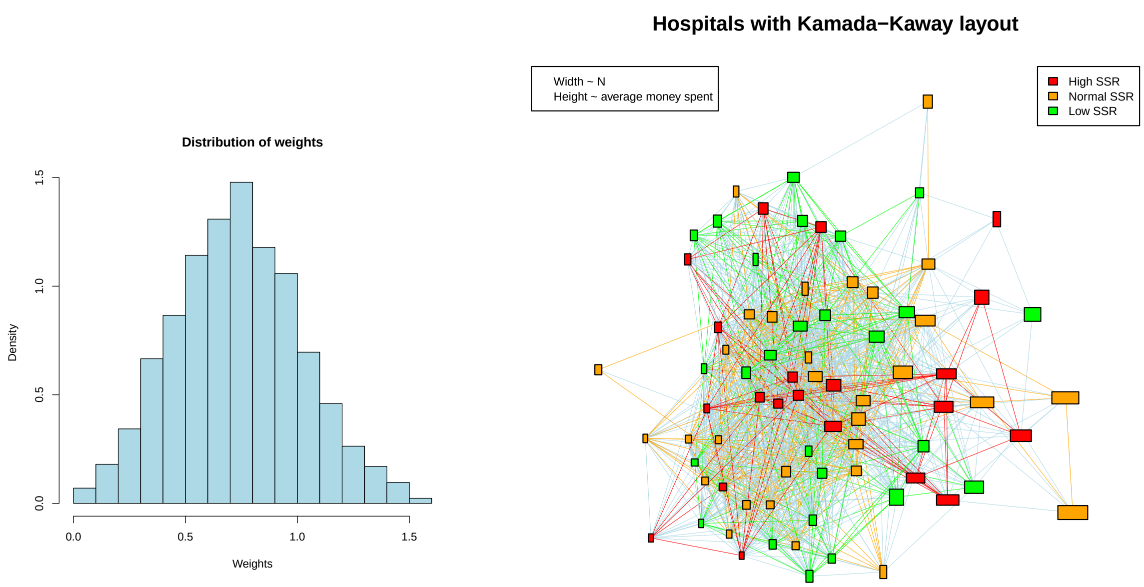


Figure 23: Weight distribution in the similarity graph obtained with distance 15 (left) and the graph obtained keeping only edges smaller than the mean, using the continuous distance (right). The coloration is done with the information derived from the funnel plot related to SSR.

Once again our ultimate aim is to understand if similar hospitals have similar outputs and once again we must choose a cut off, so to choose the maximum distance that to similar hospital can have. The more natural cut off is to chose edges bigger than the mean or than the median. This network is still difficult to understand, but the similarity between hospitals are more even. In

Figure 23 we see the graph that derives from such cut off. We notice an interesting grouping of red hospitals in the middle of the graph, with five "red" hospitals that are really near to each other. The same grouping algorithm is applyed to evaluate if the clustering have some relation with hospital performance and in Figure 24 the results are shown.

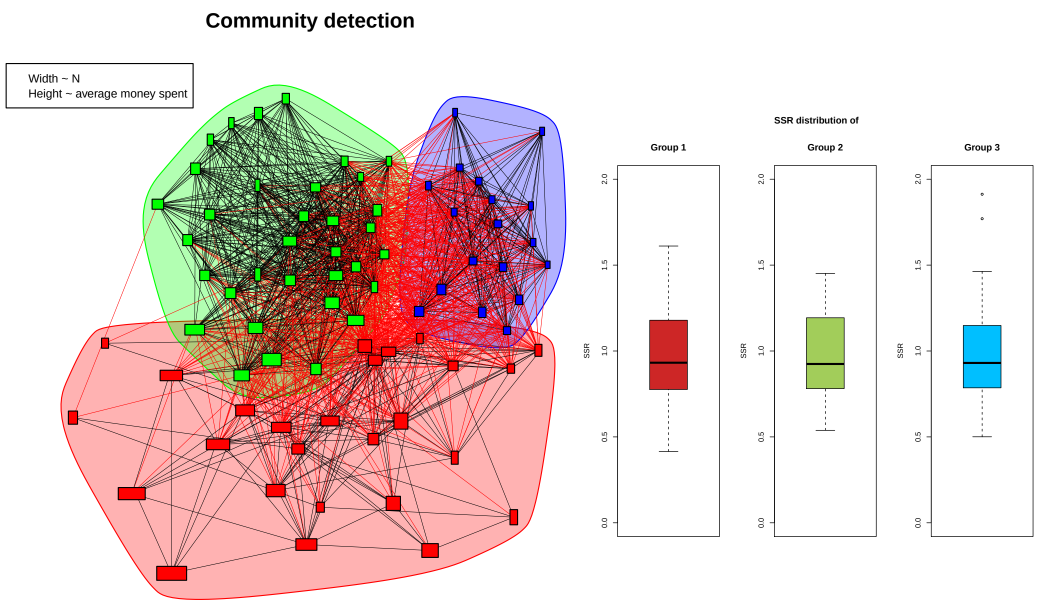


Figure 24: Community detection in the Graph in Figure 25 (left) and distribution of SSR stratified by groups found with Community Detection Algorithm fastgredy.community (right).

If we analyze the three groups in relation to all the vertex characteristics (see Table 14), we find no differences in the performance indices such as SSR and re-hospitalization, as we can clearly see in Figure 24 for the SSR. The *DRG* flag results significantly different between the three groups, but that brings no surprise, since the *DRG* flag is strongly related with the mean worth and the percentage of *Cardiochirurgy events*, both used in the evaluation of the edges. A surprise is instead the difference between the patients risk that we see in Figure 25 and it's interesting the concurrence in group 3 of low mean worth and high patient risk.

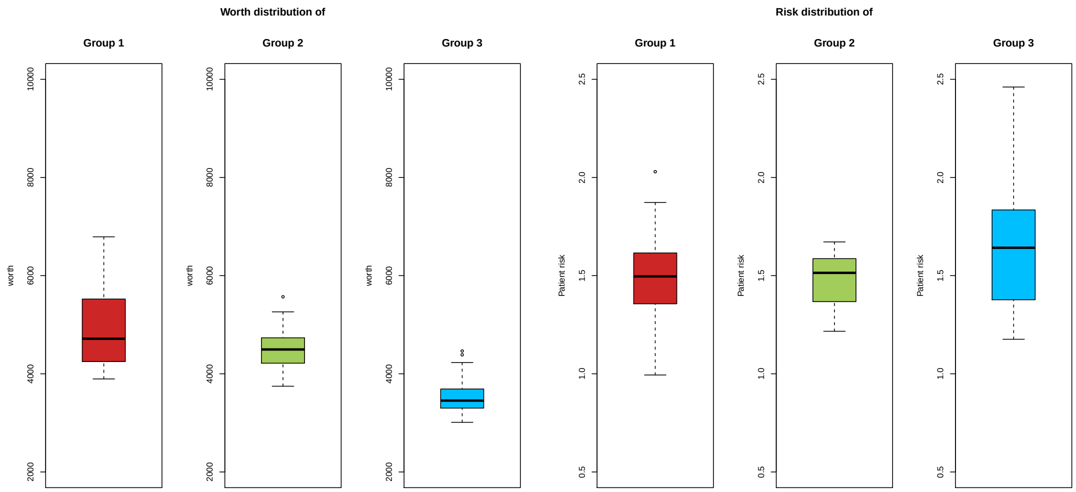


Figure 25: Distribution of average worth (left) and average risk (right) stratified by groups found with Community Detection Algorithm fastgredy.community.

8 Statistical models for network analysis

In this section we exploit proper statistical models to befitted on the hospital networs introduced in Section 7.2, i.e., *Gravity Model* and other different generalized linear models.

8.1 The hospital network as a flow graph

The first network we examine is the flow graph presented in Section 7.2, a network where edges represent the patients flow. In literature this flow is usually named as "traffic". Flows are our main interest: we'd like to understand the flow as a response variable and to explain it's behaviour looking at vertices and edges features.
Given the network $G = (V, E)$ we can identify a matrix that sum up all the responses: the $Origin - Destination$ (OD) matrix. Given any couple of vertices i and j, the OD matrix is defined as $\mathbf{Z} = Z_{ij}$, where Z_{ij} is the volume of patients flowing from hospital i to hospital j.
Hereafter, in all the models fitted on the flow network $\mathbf{Z}$ will represent the response matrix

Gravity models

Gravity models are a class of models that describe the interaction between different groups or population. They have been used mainly in geography, economics and sociology. The term "gravity model" derives from the fact that the dimension of the two populations and the distance (real, or metaphorical) between the two are considered as explanatory variable.
This kind of modeling consider all edges as indipendent between each other so, while it can be usefull in explaining the phenomena to some extent, it must be used with caution. The independence assumption is a problematic issue, because edges with a vertex in common may be dependent.
The Z_{ij} are counts, we assume Z_{ij} with independent Poisson distributions with mean:

$$E(Z_{ij}) = h_O(i) \cdot h_D(j) \cdot h_S(\mathbf{c}_{ij}) \tag{17}$$

where h_O, h_D and h_S are positive functions of the origin i, destination j and a vector $\mathbf{c}_{ij}$ of k separation attributes. The k elements of $\mathbf{c}_{ij}$ are chosen to describe some kind of distance or cost.
A classical example is the model of Stewart, developed for demograpy, in which:

$$E(Z_{ij}) = \gamma \cdot \pi_{O,i} \cdot \pi_{D,j} \cdot d_{ij}^{-2} \tag{18}$$

Where π is a measure of dimension of the origin or destination and d_{ij} is a measure of the distance. A generalization usually used is:

$$E(Z_{ij}) = \gamma \cdot \pi_{O,i}^{\alpha} \cdot \pi_{D,j}^{\beta} \cdot c_{ij}^{\gamma} \tag{19}$$

We use the form proposed in (19) to fit two simple models.

Model 1.1: $\log E(Z_{ij}) = K + \alpha \log SSR_i + \beta \log SSR_j + \gamma \log d_{ij}$

Model 1.2: $\log E(Z_{ij}) = K + \alpha \log N_i + \beta \log N_j + \gamma \log d_{ij}$
Where d_{ij} is the distance identified in (13) between hospital i and hospital j.
As Goodness Of Fit (GOF) indicators we use the values of AIC, BIC and Loglikelihood. We also look at the plot of fitted against measured flows.
In Model 1.1, despite coeffcents are significant, the values of the parameters are really small and the GOF very poor, so we don't show the results. On the other hand in Model 1.2 we find that hospital's dimension is significant in assessing the dimension of the edge, and the GOF of the model improved. The corresponding results are reported in Table 19. 18.

	Model 1.2
(Intercept)	-12.60^{***}
	(0.13)
lN_in	1.81^{***}
	(0.03)
lN_out	2.11^{***}
	(0.02)
ldistance	0.13^{***}
	(0.01)
AIC	161,364.65
BIC	161,391.66
Log Likelihood	-80,678.33
Deviance	150,892.50
Num. obs.	6,320

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

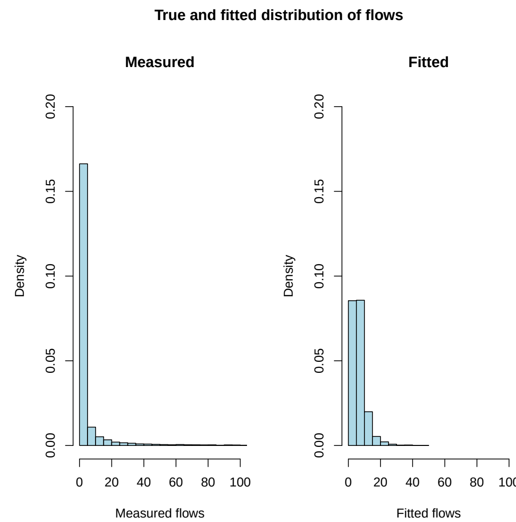


Table 18: Summary of fitted Model 1.2 (left) and histograms of measured versus fitted for Model 1.2 (right)

First of all we see that results are coherent with the remarks we made before. The coefficent of N_{in} is postitive, since a great number of patients means a greater number of chronic patients, that are responsible for the traffic between hospitals. Also the coefficient of N_{out} is positive, therefore the "head" hospital dimension increase the flow. The coefficient relative to the difference between hospitals is positive, meaning that distance between hospital increases the flow, but we see that the value of the coefficient is very small, and the distances between hospitals are in most cases smaller than one, so this dissimilarity beteen structures isn't really a major factor.

The problem is that yet again the goodness of fit is not very strong: the comparison between the measured values and the fitted values are poor. The strong presence of latent factors probably hinders our capacity in modeling the problem.

Further models

After the gravity models, we use all the results obtained in the descriptive analysis, to point out the right variables to explain edges dimension. The Z_{ij} is once again in the form of counts, with independent Poisson distributions and mean functions $E(Z_{ij})$.

Some among the variables in the Table 19 are used in the following models

Variable	Description
N	Hospital size
$group$	group label of the hospital, derived from graph with edges >50
$group_{equal}$	1 if $group_i$ and $group_j$ are equal, 0 if not
reh	percentage of re-hospitalization (first admission)
$group_i : group_j$	interaction between groups
$hosp$	hospital label

Table 19: List and explanation of variables used in the analysis.

Same group effect - Model 2.1:

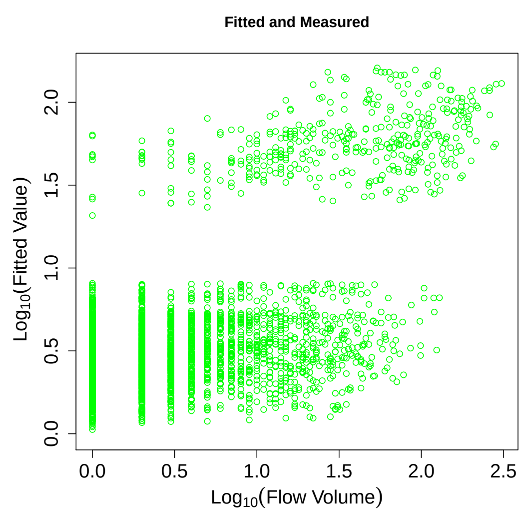
$$\log E(Z_{ij}) = a \log N_i + b \log N_j + \alpha(|N_j - N_i|) + \beta(group_{equal}) + \gamma reh_j + \phi reh_i$$

$$E(Z_{ij}) = K \cdot N_i^a \cdot N_j^b \cdot e^{\alpha(|N_j - N_i|) + \gamma reh_j + \phi \cdot reh_i} \quad (20)$$

Firstly we use the hospital sizes as suggested by the Model 1.1. The choice of the covariate $|N_j - N_i|$ is based on the idea we had in Section 7.2, that the difference between hospitals size increase the flow between hospitals. In the term $group_{equal}$ we use the grouping we have found using the graph with edges bigger than 50, since hospitals in the same area are more likely to have flows between them. The term reh_i is adopted since an hospital that produce more rehospitalization has more probability to produce flows. We introduce also the term reh_j. The results are reported in Table 20.

	Model 2.1
(Intercept)	-8.71^{***} (0.16)
lN_in	1.10^{***} (0.03)
lN_out	1.36^{***} (0.03)
dif_N	0.00^{***} (0.00)
g_uguale1	2.96^{***} (0.01)
reh_in	0.52^{**} (0.16)
reh_out	0.79^{***} (0.17)
AIC	81,605.97
BIC	81,653.23
Log Likelihood	-40,795.98
Deviance	71,127.82
Num. obs.	6,320

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$



(a) Show the plot of fitted flows against measured flows.

Table 20: Summary of fitted Model 2.1 (left) and plot of measured versus fitted for Model 2.1 (right)

Once again the number of patients in hospitals has a positive coefficient and the same is true for the difference of hospital dimension, but the coefficent is really small. The exposure of the hospital the flow depart by has a positive coefficient, as expected, The dimension of the *out* hospital has a positive coefficient, too, meaning that patients are likely to move to big structure. The absolute difference between hospital exposures is not significant, or else, it's significant, but really small. Both Re-hospitalization rate in the hospital *out* and in the hospital *in* have positive coefficients. The coefficients relative to re-hospitalization means that if the re-hospitalization rate are high, the flow raise, meaning that if both hospitals have high hospitalization the flow is enhanced more.
In Figure (20a) we see the models has two problems: firstly the smaller edges are too many for a Poisson model and the fitted values are completely wrong for small flows, secondly while the coefficient relative to $group_{equal}$ is surely significant, there are probably groups that are more connected than others, while in Figure 20a the fitted values separates in two different groups. To solve this second problem, we propose the Model 2.2.

Grouping effects - Model 2.2:

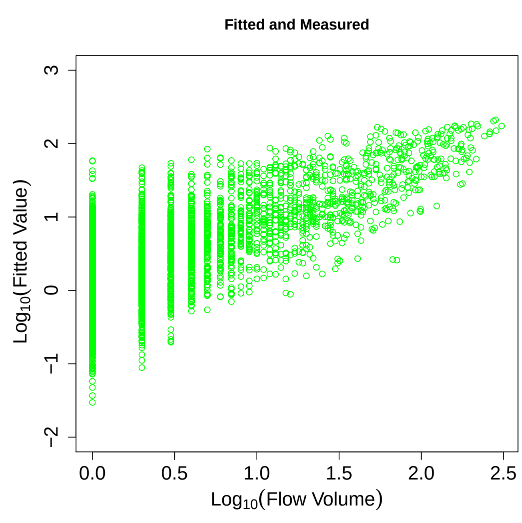
$\log E(Z_{ij}) = a \log N_i + b \log N_j + \alpha(|N_j - N_i|) + \xi(group_i) + \phi(group_j) + \gamma reh_j + \eta reh_i + \theta(group_i : group_j)$

$$E(Z_{ij}) = K \cdot N_i^a \cdot N_j^b \cdot e^{\alpha(|N_j - N_i|) + \xi(group_i) + \phi(group_j) + \gamma reh_j + \eta reh_i + \theta(group_i : group_j)} \quad (21)$$

The main difference from Model 2.1 is that instead of using the $group_{equal}$ flag, we use three variables, the group of hospital i and j and the interaction between the two of them. In Table 21 we show only some of the coefficients, but also if they aren't shown, they are for the most part significant.

	Model 2.2
(Intercept)	-12.13^{***} (0.21)
lN_in	1.62^{***} (0.03)
lN_out	1.88^{***} (0.03)
dif_N	0.00^{***} (0.00)
reh_in	3.05^{***} (0.22)
reh_out	3.46^{***} (0.22)
....	....
g_in10	-1.32^{***} (0.05)
g_in11	-1.26^{***} (0.04)
....	....
g_out10	-1.31^{***} (0.05)
g_out11	-1.36^{***} (0.05)
....	....
g_in10:g_out10	3.09^{***} (0.09)
g_in11:g_out10	-0.67^{*} (0.27)
AIC	50,341.26
BIC	52,812.30
Log Likelihood	-24,804.63
Deviance	39,145.12
Num. obs.	6,320

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$



(a) Show the plot of fitted flows against measured flows.

Table 21: Summary of fitted Model 2.2 (left) and plot of measured versus fitted for Model 2.2 (right).

Coefficients relative to same variables have similar values both in Model 2.1 and 2.2, only the coefficient relative to re-hospitalization increases significantly moving from Model 2.1 to Model 2.2. We notice that most of the groups and the interactions are significant, so that we can estimate areas where the flows are mostly outward or inward directed. Also we can look at the interaction between different groups and find significant connections between them.
BIC, AIC and Log likelihood improve and also the plot of fitted against measured flows improve significantly moving from Model 2.1 to Model 2.2. The interaction between groups erase the division within fitted values that we had in Model 2.1. The smaller edges remains randomly distributed, as in the prior model.

Hospital effects - Model 2.3:

$\log E(Z_{ij}) = a \log N_i + b \log N_j + \alpha(|N_j - N_i|) + \beta(group_{equal}) + \theta hosp_i$

$$E(Z_{ij}) = K \cdot N_i^a \cdot N_j^b \cdot e^{\alpha(|N_j - N_i|) + \beta(group_{equal}) + \theta(hosp_i)} \tag{22}$$

Lastly we propose to use the hospitals (categorical variables) as explanatory variables for the flows. This time we want to assess how hospitals can influence flows. Since we want to mantain some kind of grouping effect, but we want to reduce as possible the number of explanatory variables, we use the $group_{equal}$ flag from model 2.1.

	Model 2.3
(Intercept)	5.74^{***} (1.58)
lN_in	-3.15^{***} (0.44)
lN_out	1.54^{***} (0.03)
dif_N	0.00^{***} (0.00)
g_equal1	3.15^{***} (0.01)
hosp_in030AHQ	0.49^{***} (0.07)
hosp_in030AIO	-0.49^{***} (0.08)
....	....
AIC	76,656.30
BIC	77,216.67
Log Likelihood	-38,245.15
Deviance	66,026.15
Num. obs.	6,320

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 22: Summary of fitted Model 2.3

As before, coefficients for N_{out}, $diff_N$ and $group_{equal}$ mantain similar values compared to prior models, while the coefficient realive to N_{in} changes and becomes negative. Almost all the hospitals are significant, so that we can, for every hospital, identify if the hospital is prone to an out flow or an in flow.

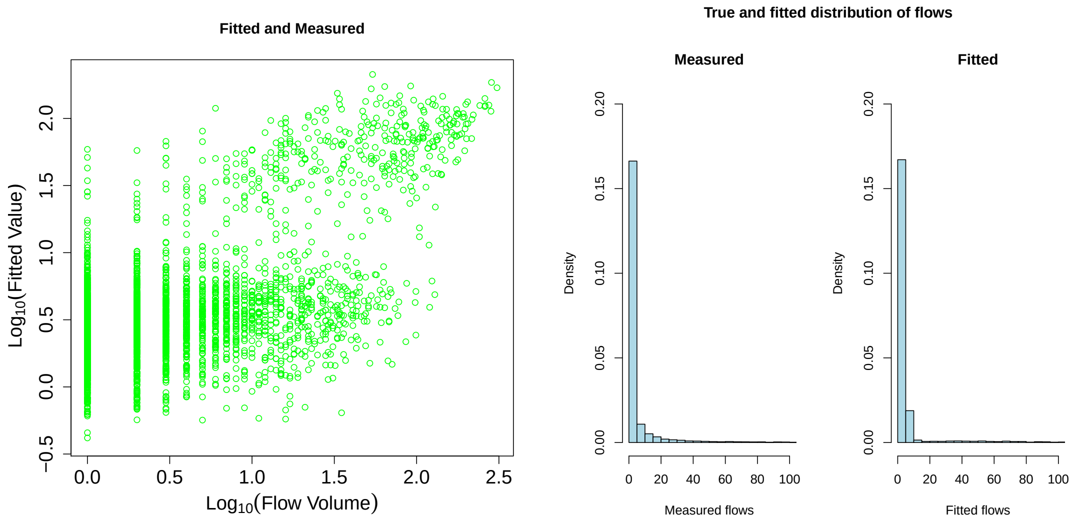


Figure 26: Plot and histograms for Measured versus Fitted for Model 2.3

AIC, BIC and Log Likelihood worsen. We see once again the group effect in the plot of fitted against measured. Fitted flows are once again divided, despite the effect is softer than Model 2.1 (see Figure (26)).

Remarks

According to the GOF analysis and the results presented above, we select Model 2.1 as the most appropriate for modelling the hospital network flow. In this model both hospitals dimensions increase the flow. While the dimension of the hospital that release the patients can be linked to the number of patients who exit form the hospitals itself, the other dimension can be a measure of the attractiveness of the structure. Also both re-hospitalization rate of *in* and *out* hospitals increase the flow. While for the hospital dimensions we can understand why re-hospitalization rate of the hospital that release the patients can be related to the flow, an explanation for the re-hospitalization rate of the hospital where the patient move is more difficult. Among group effetcs and group interaction effects coefficients are mostly significant, so we can see which groups are more prone to outflow or inflow and what groups communicate best.
We notice that this kind of graph can be really usefull for dealing with transfers. During an "HF event" a patient can be moved from an hospital to another. While in the database we analyze all the adjacent transfers are coercerd in one event, in the original database all the single hospitalizations are separate rows. With this approach, applyed on the original data, we can identify hospitals that are in-stars and out-stars for transfers and also analyze flows that clinicians can identify as not common (such as transfer from an hospital with cardiology to an hospital without). At the same time an analysis on such a graph can give usefull insights on the nature of transfers and help with logistic issues that can arise.

8.2 GLM with $p*$ model

One of the main focus of this part is to fit predictive models for the graph seen as a whole, with edges (relational ties) as a response. Basically we want to model the relation between a similarity tie and a specific collection of explanatory variables, such as those in Table 14, considering also the dependence between the relation and the structure of the entire graph. The explanatory variables can be of several different types, taking in account the difference between nodes or their similarity, but also the graph characteristics as a whole (i.e: number of ties, number of mutual edges, the degree of the actor).
A social relation (i.e. a similarity between two hospitals) is defined on a set of social actors (i.e hospitals) and measures how these actors are linked to each other. The main statistical focus in the literature has been on models for univariate and dichotomous relations represented as directed graphs, so we need to modify our network. As previously stated choose a cut off and we put to zero all edges bigger than the cut off and to 1 all the others, so that where there is an edge there is similarity, where there isn't the two structures are considered different. On such a graph we procede with the modelling.

The $p*$ model

All similarities between hospitals can be represented by a sociomatrix, $X = [X_{ij}]$. In this matrix X_{ij} are the entries and X_{ij} is equal 1 if there is a relation between unit i and unit j, 0 if there isn't. From X, we define X_{ij}^{+} as the sociomatrix where there is always a tie from i to j, X_{ij}^{-} as the sociomatrix where there is never a relational tie from i to j and X_{ij}^{C} as the matrix that has no information about a tie between i and j.
Our response variable is the presence or absence of a tie that is specified in X_{ij}, while the explanatory variables are node covariates (such as mean expense, presence of cardiochirurgy, mean age..) or any graph-theoretic characteristic of the relation (number of ties, dimension of stars, number of triangles..)
We denote these explanatory variables by $z_1(x)$, $z_2(x), ...,$ and the model parameters, the k elements of the vector θ, are coefficients of a linear function of these explanatory variables as in standard linear models:

$$\theta_1 z_1(x) + \theta_2 z_2(x) + ... + \theta_r z_r(x) \tag{23}$$

We need to model the probability of the observed x, $P(X = x)$; but we models a logarithmic transformation of it, we then say that:

$$log[P(X = x)] \sim \theta_1 z_1(x) + \theta_2 z_2(x) + ... + \theta_r z_r(x) \tag{24}$$

In eq (24) the θ parameters are the unknown regression coefficients and must be estimated.
A great problem of this exponential model is to normalize the right side of 24, as the normalizing costant is of difficult evaluation and can be evaluated only on the simplest scenarios.
An alternative version of (24) is a logit formulation, that was first described by Strauss and Ikeda [32]. Hence we model:

$$\frac{P(X_{ij} = 1|X_{ij}^{C})}{P(X_{ij} = 0|X_{ij}^{C})} = exp(\quad \theta\, (\mathbf{z}(x_{ij}^{+}) - \mathbf{z}(x_{ij}^{-})\,) \quad) \tag{25}$$

The coefficient θ can be interpreted as the log-odds of an individual tie conditional on all others.
We notice that ERGM are designed in direct analogy to the GLMs. The idea is to facilitate the extention of well-established statistical principles and methods for the construction, fitting and comparison of models. While this class of models has potential, much of the standard inferential infrastructure available for GLM, resting on asymptotics approximations has not already been formally justified. As such, $p - values$ and other results must be used with some care.

Lastly one distinction in model terms is worth reminding: terms are either dyad independent or dyad dependent. Dyad independent terms (like nodal homophily or difference between nodes terms) imply no dependence between dyads beacause the presence or absence of a tie does not depend on the state of other ties. Dyad dependent terms (like degree terms, or triad terms) imply dependence between dyads. In standards settings, with i.i.d. distributed realizations, exponentially family models like (24) are generally fitted using the maximum likelihood method. In the context of the ERGM, the estimators of the parameters are well defined, but their calculation is non-trivial in all settings but the simplest ones. Therefore, models with dyad dependent terms are fit using the function *ergm*, which implements a version of Markov Chain Monte Carlo (*MCMC*) maximum likelihood estimation (see [17] for additional details and references).

The *ERGM* package

The *ERGM* package is developed for R and it's suggested in the statistical study of networks. It's aim is to propose tools for the modeling of network graphs and it's based on the class of models called exponential-family random graph models (ERGMs) or $p*$ models. Within this framework, one can choose between some already implemented graph statistics as explanatory variables and obtain maximum-likehood estimates for the related parameters. Moreover one can also introduce node features as response variables and comparison between node features at edges' ends, such as a matching or absolute difference, and obtain the relative estimates. In the package are also available tests for goodness-of-fit. Model comparison are allowed and the possibility to simulate additional networks (with the underlying probability distribution previously evaluated) allows tests on the goodness of the model proposed.

GOF for the MCMC: With the results of the MCMC algorithm, one can look at all the MCMC characteristics, such as autocorrelation against lags, autocorrelation between explanatory variables, accepted sample percentage, etc. Figure 27 represents the plot related to the mixing of the MCMC and is already imbedded in the package. For all the other parameters, one can look at the mcmc.diagnostic that is proposed in the package.

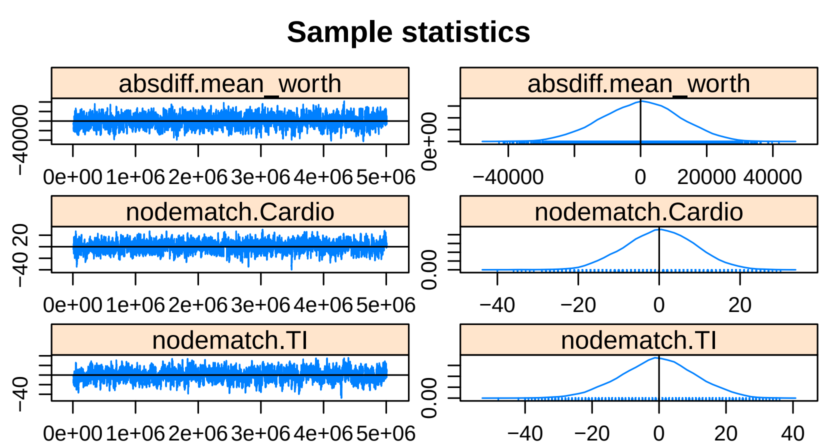


Figure 27: Convergence analysis for a MCMC.

GOF for the graph itself: A maximum likelihood estimator, while providing the best choice from the class of models we are considering, does not necessarily result in a good model. In assessing if the model is a good one, we can use the GOF function provided in the package. The idea is to compare the observed graph with a set of simulated networks (simulated with the probability distribution we have obtained) based on certain network statistics. A number of graphs are then simulated and some network statistics are evaluated for every network, these simulated network statistics are then compared with the statistics evaluated on the observed network. The sets of statistics used for the comparison are:

1. The geodesic distance distribution: proportion of pairs of nodes whose shortest connecting path is of length k. Pairs of nodes that are not connected are classified as $k = \infty$
2. The edgewise shared partner distribution: number of edges in y between two nodes that share exactly i neighbors in common, i.e., the number of edges that serve as the common base for exactly i distinct triangles.
3. The degree distribution: number of individuals with exactly k relationships.

As these characteristics are not used as parameters in the fitting of the model, an index of goodnes of fit can be the similarity between observed and simulated on these network statistics. If the observed network is too different from the simulated one, the model is not a good representation of our network and the class of models may be changed, depending on the aim of the modelling.
The plots are such as Figure 28, where the observed is the black line and the boxplots are obtained from the simulated graph. A model that is a good representation of the phenomena has at least all the black lines inside the grey ones.

Generalized Linear Models

In the fitting of the model some difficulty may arise. The fitting of a model on a nearly complete or nearly empty network may fail [3, 16]. The estimation process can be affected also when the proposed model is a very poor representation of the observed network. In the first scenario the model usually fails. If this is due to the poor representation of the graph we can try to enrich the algorithm with more variables that are more significant for the model, such as, sometimes, the same variables we used to build the graph itself. Therefore we won't study all the hospital network we have previously proposed.
The general model we use is:

$$\log(X = x|Y = y) \sim \quad \theta_1 S_1(x) + \theta_2(x) AKT_\lambda(x) + \beta^{\mathbf{T}} \mathbf{g}(x, y) \tag{26}$$

Where:

- $X = [X_{ij}]$ is the random adjacency matrix fot the graph G.
- x is the particular realization of X.
- $S_1(x) = \sum_{ij} x_{ij}$ is the number of edges.
- Y are the random attribute statistics.
- y is the particular realization of Y.
- $AKT_\lambda(x) = 3T_1 + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(x)}{\lambda^{k-1}}$ where
  - T_k is the number of k-triangles.
- $\mathbf{g}(x, y) = \sum_{1 \leq i < j \leq N_v} x_{ij} h(\mathbf{y_i}, \mathbf{y_j})$ is a function of the observed edge distribution and of the features distribution. h can be:
  - The absolute difference for a feature between the two hospitals at the edge's ends.
  - The match for a feature between the two hospitals at the edge's ends.

Model 1

Due to the nearly completeness (see [16] for a deeper discussion on relationship between completenss and degeneracy) the graph represented in Figure 19 presents fitting problems. In order to fit a feasible model mantanining as much as possible of the observed features, we choose to trucate edges which are greater than 1.01. We therefore analyze a graph that is a blending between the graph with the cut off chosen at 1 and the graph with cut off chosen at 1.4, with different groups connected by some edges.
We model the graph firstly with all variables but the dichotomous flags we used in the evaluation of distance, because we know these variables are most of the explanation for the presence of absence of the edges.
The model we choose is

$$h(\mathbf{x_i}, \mathbf{x_j}) = (|Shock_i - Shock_j|, |SSR_i - SSR_j|, |weight_i - weight_j|, |ptca_i - ptca_j|,$$

$$|icd_i - icd_j|, |cabg_i - cabg_j|, |reh_i reh_j|, |\%IT_i - \%IT_j|, worth_i - worth_j|, |N_i - N_j|,$$

$$LOS_i = LOS_j)$$

The mixing of the Markov chains is good, while the GOF of the model is somewhat lacking. In Figure 28, 30 and 29 we present the GOF results.

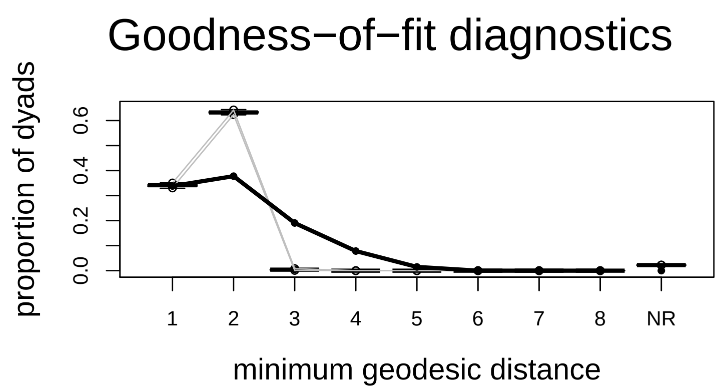


Figure 28: Comparison between simulated and observed network minimum geodesic distance.

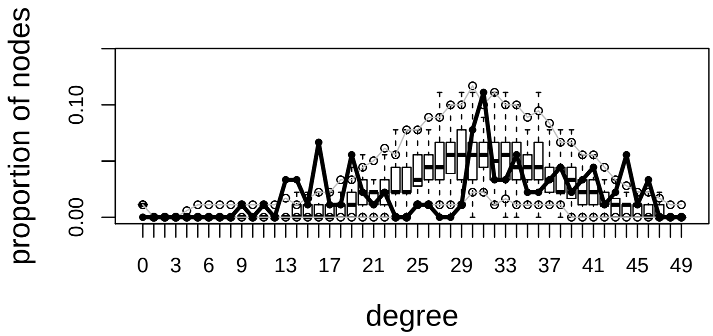


Figure 29: Comparison between simulated and observed network degree.

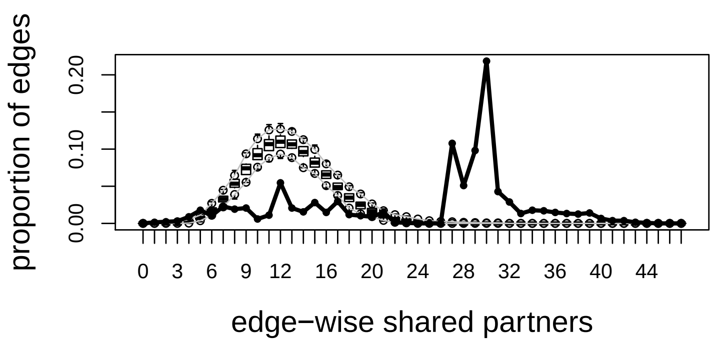


Figure 30: Comparison between simulated and observed network edge-wise shared partners.

We see that the only graph statistic that simulated networks and observed network have similar is the node degree distribution. The two othe graph statistics are really different. The lacking in the GOF is probably due to the fact that our graph is strongly dependent on the flag variables, while dependence on the structure is really low. If we fit a model with all vertices features as explanatory variables the GOF improve significantly and the simulated networks are very similar to the observed one.
In Table 23 are reported coefficents for the explanation variables we choose in the modelling.

Model with discrete distance			
	Estimate	Std. Error	p-value
edges	-12.50	1.92	$< 1e-04^{***}$
absdiff.Shock	2.87	3.07	0.35
absdiff.SSR	-0.21	0.14	0.12
absdiff.weight	-0.53	0.19	0.00561^{**}
absdiff.reh	0.97	0.93	0.29
absdiff.icd	18.31	2.71	$< 1e-04^{***}$
absdiff.cabg	52.30	9.90	$< 1e-04^{***}$
absdiff.ptca	-7.40	1.75	$< 1e-04^{***}$
absdiff.p it	-2.20	0.41	$< 1e-04^{***}$
absdiff.worth	-0.001	$9.69e-05$	$< 1e-04^{***}$
nodematch.los	0.28	0.092	0.00252^{**}
absdiff.Np	$-2.79e-04$	$2.26e-05$	$< 1e-04^{***}$
AKT	4.65	0.66	$< 1e-04^{***}$
AIC	4,267		
BIC	4,348		

Table 23: Coefficients for the model evaluated on the graph obtained with discrete distance keeping only edges smaller than 1.01. The model is fitted with all the variables available for vertices, except flags for Cardiochirurgy, Emodynamic and IT care.

Coefficients relative to edges have a negative value, so that the probability that a tie is produced is very low unless some vertex features are assumed.That is coherent with the nature of the graph, that is nodes exist only if related features affirm so. Indeed, the similarity for *ptca* and *IT* percentage replaces Emodinamics and Intensive therapy flags, trying to explain the similarity between flags and increasing the probability of an edge where the difference is small. More difficult to understand are the ICD and CABG percentage, that with high values that boost the probability of an edge.
The AKT factor has a positive coefficient and aknowledges the presence of strong transitivity effects. We fit the model also with the GWD, but the related coefficinet is not significant, probably due to the strong dependence on the graph to node covariates.
SSR, Shock events percentage and re-hospitalization rate are never significant in explaining the presence or absence of an edge and that is also coherent with the tests we made in Section 7.3. Performance indices such as Standardized Mortality and re hospitalization rate don't seem to be related to the similarity we choose.

Model 2

Due to the polarization issues in the distance as defined in (13), we analyzed also the graph built with distance as specified in equation (15). We show only results for the complete model, since without all the variables the model doesn't converge. The continue distance has a weight distribution as shown in Figure 23 and we need to choose a reasonable cut off. Since we must choose a value for which all edges smaller than the cut off state a similarity relation and the weight distribution is simmetrical around the mean, we choose the mean and the first quartile and we show results for both choices.
The model is as equation (26) with

$$h(\mathbf{x_i}, \mathbf{x_j}) = (|Shock_i - Shock_j|, |SSR_i - SSR_j|, |weight_i - weight_j|,$$

$$, |ptca_i - ptca_j|, |\%card_i - \%card_j|, |icd_i - icd_j|, |cabg_i - cabg_j|,$$

$$|reh_i reh_j|, |\%IT_i - \%IT_j|, worth_i - worth_j|, |N_i - N_j|, LOS_i = LOS_j)$$

As before, the MCMC converges and both the mixing of the chain and the autocorrelation are acceptable. As for the goodness of fit analysis of the resulting graph was acceptable too. Coefficients relative to the model with the cut off chosen as the mean are in Table 24, the ones relative to the model with the cut off chosen as the 1^{st} quartile are in Table 25.

Model with all variables,	continuous distance	and cut off = mean	
	Estimate	Std. Error	p-value
edges	$1.093e+01$	$1.087e+00$	$< 1e-04^{***}$
absdiff.Shock	$-7.388e+00$	$5.498e+00$	0.1791
absdiff.SSR	$-1.263e+00$	$2.733e-01$	$< 1e-04^{***}$
absdiff.worth	$-1.159e-04$	$1.532e-04$	0.4491
absdiff.weight	$-4.982e-01$	$2.973e-01$	0.0939.
absdiff.ptca	$-7.926e+01$	$4.421e+00$	$< 1e-04^{***}$
absdiff.icd	$-4.860e+00$	$4.377e+00$	0.2669
absdiff.IT	$-3.139e+01$	$1.418e+00$	$< 1e-04^{***}$
absdiff.cabg	$-1.346e+02$	$3.154e+01$	$< 1e-04^{***}$
absdiff.reh	$2.503e-01$	$2.607e+00$	0.9235
absdiff.card	$-1.091e+02$	$5.832e+00$	$< 1e-04^{***}$
nodematch.los	$1.104e+00$	$1.692e-01$	$< 1e-04^{***}$
absdiff.Np	$-1.034e-03$	$4.829e-05$	$< 1e-04^{***}$
AKT	$-2.617e-01$	$3.465e-01$	0.4501
AIC	$1,666$		
BIC	$1,750$		

Table 24: Coefficients for the model evaluated on the graph obtained with continuous distance keeping only edges smaller than the mean of weight distribution.

Model with all variables,	continuous distance	and cut off = 1^{st} qt.	
	Estimate	Std. Error	p-value
edges	$3.845e+00$	$7.493e-01$	$< 1e-04^{***}$
absdiff.Shock	$2.421e+00$	$5.930e+00$	0.683055
absdiff.SSR	$-1.219e+00$	$3.264e-01$	0.000192^{***}
absdiff.worth	$-6.570e-04$	$2.111e-04$	0.001874^{**}
absdiff.weight	$-1.443e+00$	$4.011e-01$	0.000325^{***}
absdiff.ptca	$-7.848e+01$	$6.109e+00$	$< 1e-04^{***}$
absdiff.icd	$1.080e+01$	$5.229e+00$	0.039019^{*}
absdiff.IT	$-3.167e+01$	$1.662e+00$	$< 1e-04^{***}$
absdiff.cabg	$-1.589e+02$	$5.376e+01$	0.003135^{**}
absdiff.reh	$-2.034e+00$	$3.335e+00$	0.541938
absdiff.card	$-1.070e+02$	$7.130e+00$	$< 1e-04^{***}$
nodematch.los	$1.308e+00$	$1.931e-01$	$< 1e-04^{***}$
absdiff.Np	$-9.722e-04$	$6.345e-05$	$< 1e-04^{***}$
AKT	$1.074e+00$	$2.052e-01$	$< 1e-04^{***}$
AIC	$1,155$		
BIC	$1,240$		

Table 25: Coefficients for the model evaluated on the graph obtained with continuous distance keeping only edges smaller than the first quartile of weight distribution.

Results are similar as for model 8.2 on page 63 as the rehospitalization and the shock percentage aren't significant in assessing the presence of an edge. The SSR instead seem to have some kind of relation with the presence or absence of an edge. As the coefficient is negative two hospital with different SSR are not so likely to have an edge between the two of them as two that have similar SSR. That is an interesting concept: while the similarity in the graph with the discrete distance depends mainly with the presence or absence of a structure, the continuous distance assesses mainly similarity of treatment. So if two hospital are similar, so they have more or less same size, but also similar policies, they are less likely to have different mortality. We notice that in the second model the AKT coefficient is not significant, once again that is probably due to the relative independence between dyads.

Both models, if different, are useful in the analysis. The first model highlight yet again the apparent independence beetwen SSR and Re-hospitalization rate and the similarity introduced with distance (13) and this suggests to look at models like the second for further analysis. The second model indeed look at the similarity between hospitals differently. With this model we find that a similarity in percentage of procedures is linked with alike result in SSR. Therefore we could maybe try to understand why and how these procedures change performances.

9 Concluding remarks and future work

In this work we have examined the hospitals in Lombardia firstly as units, than as a network. As for the analysis with funnel plots we can say that the funnel plots are very useful to assess hospital quality. Evaluation of results for hospitals is of great importance and funnel plots, while avoiding a spurios ranking, point out structures that behave as outliers, so that they can be inspected further. While funnel plots seem to be the right approach, both the performance indices we choose presented overdispersion. Therefore, if institutions outside the funnel are not all outliers, a correction for overdispersion is needed. In evaluating the hospitals a decision must be made: what overdispersion correction to choose, so to be more lenient with smaller structures or with bigger ones. Indeed the choice of the overdispersion correction must be discussed throughoutly because, as we have seen, the hospitals that are identifed as outliers changes completely. Another remarks that we have made is that overdispersion lessens with the number of hospitalization and that hospitals that are outliers at third admission, are outliers also for the first. While this is interesting, we don't think that this can be of any help in classifying hospitals, since in a plausible period of time, like one year, only very bis hospitals can be evaluted for third hospitalization.

Our analysis of hospital performances have dealt with mortality and re-hospitalization, but evaluating hospital performance remains a difficult and complex task. We thereby suggest to evaluate at least two other performance indices other than pure in-hospital mortality and re-hospitalization [5]. Mortality within 30 days after the discharge is clearly an outcome of interest, underperforming beeing a high mortality rate within a certain amount of days from the dimission. The re-hospitalization within a certain period is also of interest: HF is a chronic disease, but the re-admission in a short time after the discharge is an underperformance. Hospitals can therefore be tested on: in-hospital mortality, mortality at 30 days, re-hospitalization, re-hospitalization at 30 days. Also, results at patients level were the main concern in this study, however the mean espense or the LOS of patients can be similarly analyzed. If that is the output of interest, the mean espense at patient level can be easily normalized and a funnel can be used to identify hospitals that use too much resources.

Where the funnel plots are useful with a dataset that cover a long period of time, the Cumulative Sum charts are appropriate to study the progress of individual institutions over time. As the administrative dataset are constantly updated, the Cumulative Sum charts can help hospitals in monitoring their performance during the time in which data are collected for a new evaluation, as an example with funnel plots.

When we introduced the network approach, not only new kinds of outliers were introduced, such as in-star or out-star outliers, but also a new set of interesting features and a new way to look at hospitals themselves. As patients flow connects hospitals, hospitals that have too many patients that leave the hospital or too many patients that enter the hospital can be in need of a further exam. We notice that in-star or out-star outliers are not mutual with previously introduced outliers, but some insights on the local distribution of patients affected by hearth failure and their movements can be of use. With more informations about hospitals ID, we could, as an example, analyze if private and public hospitals are connected and gain some useful insights about reasons under these transfers.

The flows we have studied are related mostly to cronic patients and, while this kind of approach give interesting results, they are difficult to understand. We propose to use a similar analysis in the study of transferts, patients that, during the HF event are transferred from an hospital to another. Precious insights can be gained: on one hand understanding what structures will have more in-transfers can prepare in the dealing of transfers, on the other hand figuring out why these transfers happen can help understanding the hospital performance. As an example a high in-flow of high-risk patients can change a hospital performance for the worse.

The poisson modelling of flows, while assuming independence between flows let us verify some of the reasons behind patients flow. Among all the results we highlight the fact that hospital

performance related to mortality was not one of the main causes.
The idea of a distance between hospitals remains an interesting idea, especially since the analysis has let us know the presence/absence of some kind of structures, such as Intensive Therapy, seem to have little effects on the results on mortality. Such knowledge strongly suggests to evaluate an hospital with performances and not only presence or absence of structures. Therefore, we suggest to analyze all the data related to procedural aspects, so to understand the best course of action in the treatment of Heart Failure. Indeed, in the graph flow, we have seen that there seem to bee some areas were the results are better and with the introduction of a continuous distance, we saw that a similarity in percentage of procedures is linked with alike result in SSR, so that some procedures seem somewhat related to results. We could maybe try to understand why some clusters seem on average better than others. Also, if these areas are connected with geographical districts, maybe some differences in protocols can be further analyzed.
As for the use of the "ergm" package, we see that it can help to take in the dependence between edges and also the nature of the graph. In this work we have seen the possibilies and we have explored the package and it's uses. We suggest the ergm package to be used on the flow graph, specifically reduced as a binary directed graph. While the choosing of the cut off is as usual a sensible matter we think the flow graph could take more advantage in the use of the structure of the graph as a regressor, more than the model with the distance we have implemented. Transitivity, presence of triangles and other graph characteristics can be assessed and these effects in a flow graph, if present, have a precise meaning.
The analysis of hospital networks discussed in this work is a first step both for the study of patients flows and for the analysis of procedural similarities. A better understanding of the flow network, particularly with ergm models, can help hospitals in handling transfers. This additional information about flows and procedural similarities can also help improving hospitals evaluation.

10 Appendix: R Codes

Funnel Plot of SSR without correction and Winsorization

This code refers to the evaluation of the funnel plot of SSR with limits evaluated without overdispersion correction as in Section 2.1. The figure obtained is as in Figure 4.

```
val <- seq(0, 2000, 0.5)
p1 = 0.99
p2 = 0.01
p3 = 0.975
p4 = 0.025
theta = 1

plot(ssr_den,ylim=c(0,6), ssr, type="p",
  xlab="Expected␣cases", ylab="Observed/Expected",
  main="In-hospital␣deaths␣after␣first␣hospitalization", pch=16)
abline(h=theta)
points(val, theta+c(qnorm(p1)*sqrt(1/val)), type='l', col='red')
points(val, theta+c(qnorm(p2)*sqrt(1/val)), type='l', col='red')
points(val, theta+c(qnorm(p3)*sqrt(1/val)), type='l', col='blue')
points(val, theta+c(qnorm(p4)*sqrt(1/val)), type='l', col='blue')
```

The code refers to the Winsorization correction explained in Section 2.2. the evaluated ϕ and τ are used in the funnel limits evaluation as in (5) and (9) that are shown in Figures 5 and 4.

```
## Winsorization for SSR ##
q1 <- 0.05
q2 <- 0.1
q3 <- 0.15
theta <- 1

Z<-(SSR-theta)*sqrt(ssr_den)
Zq1=quantile(Z, q1)
Zq2=quantile(Z, 1-q1)

for(i in 1:length(Z))
{
  if(Z[i]<=Zq1)
    Z[i]<-Zq1
  if(Z[i]>=Zq2)
    Z[i]<-Zq2
}

phi1 <- sum(Z^2)/n_osp
w <- ssr_den
tau21 <- (n_osp*phi1-(n_osp-1))/(sum(w)-(sum(w^2)/sum(w)))
```

Cumulative Failure Chart

This code refers to the evaluation of the Cumulative Failure chart introduced in Section 3. The figure obtained is as in Figure 3.

```
# boundary evaluation
```

```
p0 = 0.1  % acceptable mortality rate
p1 = 0.15  % unacceptable mortality rate
alpha = 0.05
beta = 0.2
a = log( (1-beta)/alpha )
b = log( (1-alpha)/beta )
P = log(p1/p0)
Q = log( (1-p0)/(1-p1) )
s = Q / (P+Q)
acc <- s*val-b/(P+Q)
n_acc <- s*val+a/(P+Q)

# Cumulative Failure chart plot
c_1 <- cumsum(data$dec_intra[which(data$strutt_ric_ricod==HOSP)])
plot(c_1, ylim=c(0,1500), type="l", xlab="Case␣number",
  ylab="Cumulative␣failure", main=paste("Hospital␣HOSP")
points(val, acc, type='l', col='green')
points(val, n_acc, type='l', col='red')
```

Graph visualization

In the code below we decorate the graph.
Before plotting the graph with kamada-Kaway layout we:

- Create the graph from the two data frame containing edges and vertices lists.
- Set vertices form and sizes, so to show number of patients and mean expense in the plot.
- Set edges and arrow sizes.
- Select subgroups colored with SSR (or rehospitalization rate) and color the edges as the head hospital.

The figure we obtain is the one reported in Figure 11.

```
# DEFINING the GRAPH
library(sand)
hosp.graph <- graph.data.frame(EDGES,directed=T,vertices=VERTICES)


# shapes and sizes of vertices
igraph.options(vertex.shape = "rectangle")
V(hosp.graph)$size <- 0.001*V(hosp.graph)$Np
V(hosp.graph)$size2 <- 0.001*V(hosp.graph)$worth

# shapes and sizes of edges
igraph.options(vertex.label=NA, edge.arrow.size=0.0001*E(hosp.graph)$
    weight)
E(hosp.graph)$width <- 0.02*E(hosp.graph)$weight

# Subgroups selection and coloring of the edge as the "head" hospital
FR <- V(hosp.graph)[color == "red"]
FY <- V(hosp.graph)[color == "orange"]
FG <- V(hosp.graph)[color == "green"]
```

```
E(hosp.graph)[ FY %->% FR ]$color <- "red"
E(hosp.graph)[ FR %->% FR ]$color <- "red"
E(hosp.graph)[ FG %->% FR ]$color <- "red"
E(hosp.graph)[ FY %->% FG ]$color <- "green"
E(hosp.graph)[ FR %->% FG ]$color <- "green"
E(hosp.graph)[ FG %->% FG ]$color <- "green"
E(hosp.graph)[ FY %->% FY ]$color <- "orange"
E(hosp.graph)[ FR %->% FY ]$color <- "orange"
E(hosp.graph)[ FG %->% FY ]$color <- "orange"


plot(hosp.graph, layout=layout.kamada.kawai,
  main="Hospitals␣with␣Kamada-Kaway␣layout")
legend("topright", c("High␣SSR","Normal␣SSR","Low␣SSR"),
    fill=c("Red", "orange", "green"),cex=0.7, pt.cex = 0.7)
legend("topleft", c("Width␣~␣N","Height␣~␣average␣money␣spent"),
    cex=0.7, pt.cex = 0.7)
```

Vertex and edge features

In this code we show how to obtain the plot represented in Figure13.

```
#### STRENGTH PLOT

stren_in <- graph.strength(hosp.graph, mode="in")
plot(V(hosp.graph)$Np,stren_in, main="In␣strength",
  xlab= "Vertex", ylab="Strenght", col=V(hosp.graph)$color, pch=16)
```

Here we show how to select and show a subgraph.

```
# induced subgraphs
grafo.r <- induced.subgraph(hosp.graph,FR)
plot(grafo.r, layout=layout.kamada.kawai,
  vertex.label=NA, main="Hospitals␣with␣Kamada-Kaway␣layout")
```

Here we show some among graph features of interest.

```
# SOME GRAPH FEATURES

table(sapply(maximal.cliques(hosp.graph), length)) #number of maximal
    cliques
average.path.length(hosp.graph)
diameter(hosp.graph)
transitivity(hosp.graph)

edge.connectivity(hosp.graph)
is.connected(hosp.graph, mode=c("strong"))
reciprocity(hosp.graph)
```

Here we show how to detect community in a weighted and directed graph. Also we propose both the plot with coloring by community and by SSR labelling. As specified in 7.2 we set costant edge thickness.

```
# COMMUNITY DETECTION
```

```
kc <- infomap.community(hosp.graph, v.weight=NULL)

igraph.options(vertex.label=NA,edge.arrow.size=0.2)
E(hosp.graph)$width <- 0.01

plot(kc, hosp.graph, col=V(hosp.graph)$color)
plot(kc,hosp.graph, main="Community␣of␣hospitals")
legend("topleft", c("Width␣~␣N","Height␣~␣average␣money␣spent"),
  cex=0.7, pt.cex = 0.7)
```

ERGM model

Here we show the code to implement an ergm model with the ergm package.

```
library(ergm) # Will load package 'network' as well.

hosp.graph<-graph.data.frame(EDGES,directed="F",vertices=VERTICES)

A <- get.adjacency(hosp.graph)
v.attrs <- get.data.frame(hosp.graph, what="vertices")
hosp.s <- network::as.network(as.matrix(A), directed=FALSE)
# all features are assigned as:
network::set.vertex.attribute(hosp.s, "SSR", v.attrs$SSR)
#...


hosp.ergm <- formula(hosp.s ~ edge
  + absdiff("Shock")
  + absdiff("SSR")
  + absdiff("weight")
  + absdiff("ptca")
  + absdiff("icd")
  + absdiff("cabg")
  + absdiff("reh")
  + absdiff("p_it")
  + absdiff("mean_worth")
  + nodematch("los")
    + absdiff("Np")
  + gwesp(1,fixed= T)
)

hosp.ergm.fit<-ergm(hosp.ergm,verbose=T,
  control=control.ergm(MCMC.interval=500)
  )
summary.ergm(hosp.ergm.fit)
mcmc.diagnostics(hosp.ergm.fit)
gof.hosp.ergm <- gof(hosp.ergm.fit)
plot(gof.hosp.ergm)
```

Acknowledgments

This work is within the Project "Utilization of Regional Health Service databases for evaluating epidemiology, short- and medium-term outcome, and process indexes in patients hospitalized for heart failure" (http://hfdata.cefriel.it/), funded by the Italian Ministry of Health and Regione Lombardia - Healthcare division, and the FARB (Finanziamento di Ateneo per la Ricerca di Base) project "Public Management Research: Health and Education Systems Assessment", funded by Politecnico di Milano.

References

[1] R. Alexandrescu, A. Bottle, B. Jarman, and P. Aylin. Classifying hospitals as mortality outliers: Logistic versus hierarchical logistic models. *Journal Of Medical Systems Volume: 38 Issue: 5 Article Number: 29*, 2014.

[2] C. J. Anderson, S. Wasserman, and B. Crouch. A p* primer: logit models for social networks. *Social Networks Volume: 21 Issue: 1 Pages: 37-66*, 1999.

[3] S. Bhamidi, G. Bresler, and A. Sly. Mixing time of exponential random graphs. *The Annals of Applied Probability, Vol. 21, No. 6, pag 2146 - 2170*, 2011.

[4] A. Bottle and P. Aylin. Intelligent information: A national system for monitoring clinical performance. *Health services research*, 43(1p1):10–31, 2008.

[5] A. Bottle, R. D. Sanders, A. Mozid, and P. Aylin. Provider profiling models for acute coronary syndrome mortality using administrative data. *International journal of cardiology*, 168(1):338–343, 2013.

[6] C. T. Butts, M. S. Handcock, and D. R. Hunter. *network: Classes for Relational Data*. Irvine, CA, 2014. R package version 1.10.2.

[7] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[8] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, 2006.

[9] C. T. Butts S. M. Goodreau Morris M. D. R. Hunter, M. S. Handcock. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 2008.

[10] R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*.

[11] Freeman and Linton C. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[12] Y. Kalish D. Lusher G. Robins, P. Pattison. An introduction to exponential random graph (p*) models for social networks. *Social Networks Volume: 29 Issue: 2 Pages: 173-191*, 2007.

[13] J. J. Gagne, R. J. Glynn, J. Avorn, R. Levin, and S. Schneeweiss. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *Journal of clinical epidemiology*, 64(7):749–759, 2011.

[14] S. M. Goodreau, M. S. Handcock, D. R. Hunter, C. T. Butts, and M. Morris. A statnet tutorial. *Journal of Statistical Software*, 2008.

[15] O. Groene, S. Kristensen, O. Arah, C. Thompson, P. Bartels, R. Sunol, and N. Klazinga. Feasibility of using administrative data to compare hospital performance in the eu. *International Journal for Quality in Health Care*, 2014.

[16] M. S. Handcock. Assessing degeneracy in statistical models of social networks. 2003.

[17] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, P. N. Krivitsky, and M. Morris. *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project, 2014.

[18] F. Ieva, C. Mazzali, M. Maistriello, and P. Barbieri. Methodological issues of administrative use of databases tot study heart failure.

[19] Paganoni A.M. Ieva, F. Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models. *Health Care Management Science, Special Issue IMA 2013. doi: 10.1007/s10729-013-9264-9*, 2014.

[20] T. Kamada and S. Kaway. An algorithm for drawing general undirected graphs, information processing letters. 1898.

[21] E. D. Kolaczyk and G. Csardi. *Statistical Analysis of Network Data with R.* 2014.

[22] D. C. Montgomery. *Introduction to Statistical Quality Control*. John Wiley, 2004.

[23] M. E. J. Newman. Analysis of weighted networks. 2004.

[24] M. E. J. Newman. Modularity and community structure in networks. 2006.

[25] L. Noyez. Control charts, cusum techniques and funnel plots. a review of methods for monitoring performance in healthcare. *Interactive cardiovascular and thoracic surgery Volume: 9 Issue: 3 Pages: 494-9*, 2009.

[26] A. Pfadt, I. L. Cohen, V. Sudhalter, R. G. Romanczyk, and D. J. Wheeler. Applying statistical process control to clinical data: an illustration. *Journal of Applied Behavior Analysis*, 25(3):551–560, 1992.

[27] M. E. Pouw, L. M. Peelen, H. F. Lingsma, and al. Hospital standardized mortality ratio: Consequences of adjusting hospital mortality with indirect standardization. *Plos One Volume: 8 Issue: 4 Article Number: E59160*, 2013.

[28] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[29] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *PNAS 105*, 2008.

[30] D. J. Spiegelhalter. Funnel plots for comparing institutional performance. *Statistics In Medicine Volume: 24 Issue: 8 Pages: 1185-1202*, 2005.

[31] S. H. Steiner and R. J. Cook. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*, 2000.

[32] D. Strauss and M. Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212, 1990.

[33] S. M. Williams, B. R. Parry, and M. M. T. Schlup. Quality control: an application of the cusum. *BMJ*, 1992.

[34] C. Yap, M. E. Colson, and D. A. Watters. Cumulative sum techniques for surgeons: a brief review. *Anz Journal Of Surgery Volume: 77 Issue: 7 Pages: 583-586*, 2007.