Master of Public Health

Master de Santé Publique

Deciphering Hospital Networks Using Graph Theory Methods

Yuanfei Anny HUANG

Europubhealth+ 2017-2019

Location of the practicum:
METIS Department, EHESP

Academic and Professional Advisor:
Dr Nolwenn LE MEUR-ROUILLARD
(METIS Department, EHESP)

Acknowledgements and Hacknowledgements

First and foremost, I would like to thank Dr Nolwenn Le Meur, my ever-patient and caring academic and professional supervisor for all of the help that she has given me, and all of the time and energy she has invested into this thesis project. If the number of Masters dissertations undertaken by a person follows a Poisson distribution, then the chances of having a supervisor who is Nolwenn or better is exceedingly low indeed. In layman terms, I lucked out.

Given that this is a thesis about network analysis, I would like to thank my support network. I tried to draw this graphically, but then realised that this would involve putting myself in the middle, and that's not the point of public health, or of studying networks, at all. Instead, I would like to thank in words the following groups of people: my family and friends back home in Australia and around the world, my EPH and MPH classmates in Europe, the staff at both the University of Sheffield and EHESP (especially Martine Bellanger and Kristina Berkut), the Europubhealth+ support staff (especially Marion Lecoq and Marion Gaborieau), colleagues in the METIS Department and EHESP Rennes in general, my choral group Ensemble Lumina, and just pretty much anyone who has gone out of their way to help me out over the last few months!

As for the "hack" part, I have to thank Nolwenn, once again, for being single-handedly responsible for about 90% of my skills in R. As for the other 10%, I would like to thank all the people online who go out of their way to respond to coding questions on internet forums, or to spend their time creating tutorials, without ever being thanked in papers. So here's a citation for the Kateto igraph tutorial, and the R-bloggers ergm tutorial (1,2). They have been immensely helpful. I would also like to thank Lauric Ferrat, who was instrumental in carrying out the FEHAP study cited extensively in this project. Although I have not had the chance to meet him in person, his thesis and R code have been an immense source of inspiration.

I would also like to thank the GenOuest Bioinformatics Computing Facility, for helping me run some of the enormous models featured in this thesis project!

Abstract

Objective:

For many patients, an acute hospital stay is followed by being transferred to a rehabilitation hospital. For the healthcare system, the flow of patient transfers represent an important movement and use of resources within the network. In order to examine the organisational determinants of these patient transfers within France, a previous study, FEHAP, was undertaken in 2014, using unweighted exponential random graph modelling (ERGM) techniques. This project aims to extend the original analysis by using the newer weighted ERGM technique, which allows for the consideration of the number of patients transferred along each pathway between hospitals. In doing so, this project will also assess the usefulness of the weighted ERGM technique as applied to analysing inter-hospital patient transfers.

Methods:

The original data from the FEHAP study was reconstructed into regional networks with and without hospital self-transfer loops. These comprised of a total of 54,889 inter-hospital patient transfers, across the regions of Bretagne, Lorraine and Rhône-Alpes, in the year 2012. These networks were analysed firstly using unweighted ERGM techniques, and then using weighted ERGM techniques.

Results:

Results obtained through weighted ERGM techniques corroborated with those using unweighted ERGM. Firstly, they showed that the structure of each network was not random. Particularly for Bretagne and Lorraine, the *department* variable shows the strongest effect for predicting patient transfer, with all of the other variables (*legal status*, *travel time*, *MCO beds, SSR beds, no MCO beds, no SSR beds, median MCO length of stay)* also being statistically significant. As well, the effect of the *legal status* variable was stronger for networks with loops compared to those without loops, and there is also a relationship between legal status and whether the hospital only offered acute or rehabilitation care. However, five of the weighted ERGM models for Rhône-Alpes, the largest network, failed to converge. Another limitation was that the weighted ERGM algorithm was not able to provide any goodness of fit information.

Conclusion:

Weighted ERGM appears to be a promising technique for analysing patient transfer data. However, further developments may need to occur before it is used for network simulations. Despite this drawback, this project was able to extend the findings from the 2014 FEHAP study. In particular, it confirmed that the geographic department of a hospital is an important predictor for patient transfers. It also demonstrated that the legal status of a hospital is statistically significant as a predictor, contrary to the original FEHAP study findings, where it was only significant for Rhône-Alpes. However, this effect is diminished for the networks without loops. Finally, this project also demonstrates that a range of mechanisms, possibly including competition, may explain the relationship between the length of stay at acute hospitals and the likelihood of transfer.

Résumé

Analyses des réseaux hospitaliers par les méthodes issues de la théorie des graphes

Objectif: Pour beaucoup de patients, un séjour à l'hôpital de soins de courte durée est suivi d'un transfert à un hôpital de rééducation. Pour le système de santé, ces transferts représentent une mobilisation de ressources et des mouvements importants au sein du réseau. Afin d'examiner les déterminants organisationnels de ces transferts de patients en France, une étude précédente, FEHAP, a été réalisée en 2014 en utilisant des techniques de modèles à graphes aléatoires exponentiels (ERGM) non pondérées. Ce projet-ci vise à étendre l'analyse initiale en utilisant les techniques plus récentes d'ERGM pondérées, qui permettent de prendre en compte le nombre de patients transférés via chacun des arcs empruntés entre hôpitaux. Ce faisant, ce projet évaluera également l'utilité des techniques d'ERGM pondérées appliquées à l'analyse des transferts de patients entre hôpitaux.

Méthodes: Les données originales de l'étude FEHAP ont été reconstituées sous forme de réseaux d'établissements et de transferts régionaux avec et sans boucles d'auto-transfert. Elles comprenaient un total de 54 889 transferts de patients entre hôpitaux, pour les régions de Bretagne, de Lorraine et de Rhône-Alpes, en 2012. Ces réseaux ont d'abord été analysés en utilisant les techniques d'ERGM non pondérées, puis en utilisant les techniques d'ERGM pondérées.

Résultats: Les résultats obtenus par les techniques d'ERGM pondérées ont corroboré ceux obtenus par les techniques d'ERGM non pondérées. Tout d'abord, ils ont montré que la structure de chaque réseau n'était pas aléatoire. En particulier pour la Bretagne et la Lorraine, toutes les variables étaient statistiquement significatives en ce qui concerne la prévision du transfert de patient, la variable *département* montrant l'effet le plus fort. En outre, l'effet de la variable de statut juridique était plus fort pour les réseaux avec boucles comparativement à ceux sans boucles, et il existe également une relation entre le statut juridique et le fait que l'hôpital offre uniquement des soins de courte durée ou des soins de rééducation. Cependant, pour la région Rhône-Alpes constituant le plus grand réseau, cinq des modèles d'ERGM pondérés n'ont pas réussi à converger. Une limite notable était que l'algorithme d'ERGM pondérée n'était en mesure de fournir des informations concernant la validité de l'ajustement pour aucun des modèles pondérés.

Conclusion: L'ERGM pondérée semble être une technique prometteuse pour analyser les données de transfert de patients. Toutefois, il faudra peut-être procéder à d'autres développements avant de l'utiliser pour les simulations de réseau. Malgré cet inconvénient, ce projet a permis d'étendre les résultats de l'étude FEHAP de 2014. En particulier, il a confirmé que le département géographique où se situe l'hôpital est un facteur prédictif important pour les transferts de patients. Il a également démontré que le statut juridique d'un hôpital est un facteur de prévision statistiquement significatif, contrairement aux conclusions de l'étude FEHAP initiale, qui ne l'avait démontré que pour la région Rhône-Alpes. Cependant, cet effet est réduit pour les réseaux sans boucle. Enfin, ce projet démontre que divers mécanismes, y compris éventuellement la concurrence, peuvent expliquer la relation entre la durée de séjour dans les hôpitaux MCO et la probabilité de transfert.

Table of Contents

List of Abbreviations and Acronyms

	English	French
CD	Contrastive divergence	
ERGM	Exponential random graph modelling	
ESPIC		Établissements de santé privé d'intérêt collectif
FEHAP		Fédération des Établissements Hospitaliers et d'Aide à la Personne Privés Non Lucratifs
FINESS		Fichier National des Établissements Sanitaires et Sociaux
GOF	Goodness of fit	
LOS	Length of stay	
MCMC	Markov Chain Monte Carlo	
MCMLE	Monte Carlo maximum likelihood estimate	
MCO		Médecine-Chirurgie-Obstétrique
MLE	Maximum likelihood estimate	
MPLE	Maximum pseudolikelihood estimate	
PMSI		Programme de Médicalisation des Systèmes d'Information
SAE		Statistique Annuelle des Établissements de santé
SE	Standard error	
SSR		Soins de Suite et Réadaptation
VIF	Variance inflation factor	

List of Tables and Figures

I. Introduction

1.1 Context

Structure and function are inseparable in many areas of life. This is also true in healthcare. While anatomists, pathologists and physiologists have been aware of this complementarity for millennia, this lens has only recently been applied to healthcare systems (3).

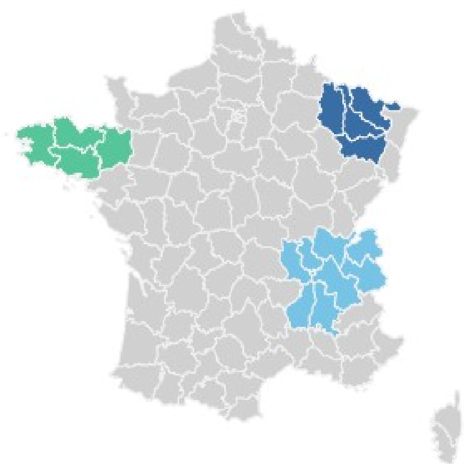
The structure of a healthcare system is, among many factors, influenced by the interaction between the actors within the system. These interactions can include both collaboration and competition (4). In turn, the function of each actor is shaped by the possibilities available within the structure of the system, including access to opportunities to collaborate. At a hospital level, actors are individual institutions, and one form of interaction between them is the transfer of patients (5). Therefore, in order to understand some of the behaviours of hospitals within the system, it is important to study inter-hospital patient transfers.

Additionally, the forces driving competition and collaboration between public and private hospitals in France have been under scrutiny in the recent years, following a series of healthcare reforms (6). Most notably, public hospital reimbursement reforms were introduced in 2008, in order to increase the competitiveness of the this sector. This has had the desired effect of increasing the market share of public hospitals in the subsequent years (7). Investigating patient transfers between hospitals is also a useful approach to examine some of the dynamics of inter-sectoral competition among hospitals.

1.1.1 Regions of France in 2012

The basis of this project is on hospital data from three regions of France, from the year 2012. At that time, there were 22 regions in France in total (8). The regions featured in this project are Bretagne, Lorraine, and Rhône-Alpes (map in Appendix x, table 1). They were chosen because of an antecedent and related study (Section 1.3)(9), against which this project will make comparisons.

Figure 1: Map of France and the regions of interest



Key:

Green = Bretagne

Dark blue = Lorraine

Light blue = Rhône-Alpes

Scale approx 1:17,000,000

Source: maps package, R software (10)

Table 1: Overview of study regions

	Bretagne	Lorraine	Rhône-Alpes
Number of departments	4	4	8
Area (km^2)	27,208	23,547	43,698
Population (2014 census)	3,341,188	2,406,226	6,500,319

Source of data: Journal Officiel de la République Française 2014 (8)

1.1.2 French hospital types and networks

Hospitals in the French healthcare system fall under the combined jurisdiction of the State, the Region and the statutory health insurance (11). There is a variety of organisational structures among the establishments, and each individual hospital may have its own characteristics, grouped under several broad categories.

Hospitals can be classified according to the type of care that they provide, and the sector to which they belong. The former classification divides hospitals into three categories (11):

- *Médecine-Chirurgie-Obstétrique* (MCO; medical, surgical and obstetric) hospitals, which are responsible for the care of acute medical conditions, and the performance of complex procedures such as surgery.
- *Soins de Suite et Réadaptation* (SSR; ongoing care and rehabilitation) hospitals, which are responsible for subacute care, such as musculoskeletal rehabilitation after surgery.
- Mixed MCO and SSR hospitals, with facilities for both acute and subacute care.

There are also three groups of hospitals according to sector (11): public, private for-profit, and *établissements de santé privé d'intérêt collectif* (ESPIC; private not-for-profit). All three sectors are able to offer both MCO and SSR care. Therefore, for the remainder of this paper, the term *statut juridique* (legal status) will be used interchangeably with hospital sector, in accordance with the terminology within the French healthcare system. As discussed above, patients are often transferred from an MCO service to an SSR service after they have completed the initial acute phase of their treatment. Patients can also be transferred from SSRs to MCOs if they once again require acute medical care. However, as discovered in the study antecedent and related to this current project, this number is significantly smaller (9).

1.2 Field of network analysis, as related to hospitals

Hospitals within a healthcare system can be conceptualised as a network (12). Networks are the combination of individual actors and the connections between them. The individuals are represented as "vertices" or "nodes", whereas the connections are termed "edges" or "links" (13). Given this, an obvious choice for studying hospital transfers is to use network analysis techniques. Network analysis techniques have been used extensively in the social sciences in recent decades (3). They are witnessing an increase in uptake in public health. This has been accompanied by an expansion in the mathematical field of graph theory, with the development of statistical techniques that allow us to progress beyond merely descriptive analysis, to hypothesis testing and being able to explore relationships in greater depth.

A recent systematic review has grouped network analysis research in public health into four broad areas of focus (14). These categories are “institutional exchange”, “physician collaboration”, “clinical co-occurrence” and “workplace interaction”. The study of inter-hospital transfers fits, unsurprisingly, into the first category of “institutional exchange”. However, among the 50 articles that fit within this category in the systematic review, none investigate the transfer of patients from an acute-care hospital to a subacute-care setting such as a rehabilitation hospital. This represents a gap in the field of study, especially in light of the fact that in three regions of France, in the context of musculoskeletal system and neurological rehabilitation alone, there were 55,196 patient transfers from an acute service to a subacute service in the year 2012 (15).

1.2.1 Definition of general network concepts

In addition to the usual measures of central tendency, which can still be applied to node and edge properties, there are additional statistics that can be used to describe a network as a whole. A review study from 2018 identified approximately 180 network statistics that have been used in studies of patient transfer networks alone (16). Out of the most commonly used of these statistics, those that are applicable to the project are outlined in Appendix 1.

Furthermore, there are a number of terms that are widely used to describe the features of all networks. Particularly relevant to this project is the concept of a directed graph. This is a network whose edges have a distinct origin and destination (13). In the networks for this project, the origin is the hospital providing acute care, and the destination is the hospital providing subacute care. An edge travelling away from a particular node is known as an out-edge for that particular node, the sender node. When it arrives at another node, it is referred to as an in-edge for the receiver node. One type of a directed edge is a loop. This is an edge that exits from and re-enters the same node. In this project, loops occur in hospitals that are able to offer both acute and subacute care, so that when patients complete the acute phase of their medical care, they are transferred to the subacute part of the same hospital for rehabilitation. Typically, in diagrams of directed networks, arrows are added to edges to indicate direction.

Another important feature of a network is whether it is weighted (13). A weighted network is where a particular attribute of each edge can vary in magnitude. For example, the networks in this project are weighted by the number of patient transfers along each edge. This means that the edges along which more patients are transferred would have a higher weight than edges with low patient transfer activity.

Finally, networks can also be considered as unipartite or bipartite (17,18). A unipartite network contains only one type of nodes. A bipartite network features nodes of two types. For example, in a healthcare context, there may be nodes representing doctors, and nodes representing patients.

Typically, in a bipartite network, there are also restrictions in terms of the links that are allowable between the two types of nodes. For example, that nodes of the doctor type can only form links with nodes of the patient type, and vice versa. By extension, multipartite networks, involving more than two types of nodes, are also possible. The decision to keep the project networks as unipartite will be detailed in the Discussion section.

1.2.2 Exponential random graph modelling

Exponential random graph modelling (ERGM) is a technique that has emerged expressly for the study of networks, where the observations are highly linked (19). Specifically, it seeks to answer the question of: "given that the actors within this network have a predisposition to establish connections in this particular way, what is the probability that in this network that one is observing, these specific links exist?" The null hypothesis is a random network, against which the observed network is compared. Other techniques that have been used in the literature to investigate hospital networks include assessments of bivariate correlations and regression analysis using a generalised linear model such as logistic regression or negative binomial regression (14). An obvious impediment to these approaches is that network data violate the assumption of independence of observations (19). Compared to these techniques, ERGM has the advantage of not having this particular requirement for independence. On the contrary, in fact, it specifically seeks to study the dependence between actors.

However, one of the major drawbacks of earlier ERGM techniques is that the existence of an edge linking two nodes is treated as a binary outcome (20). In other words, it gives yes-no answers, where two actors in the network are either linked, or they are not. The magnitude of this link cannot be considered. This is the reason behind the development and introduction of the weighted ERGM technique in 2012. With weighted ERGM, the weight of the link can be considered, if this weight can be expressed as a count variable. An example of the weight of a link being a count variable would be the number of patients transferred from one hospital to another. This is to say that one can now expand on the question that one can ask the model, so that it is now: "given that the actors within this network have a predisposition to establish connections in this particular way, what is the probability that in this network that one is observing, these specific links exist with these particular weights?" After the publication of the mathematical derivation for weighted ERGM, the relevant software code able to manage these calculations needed to be developed and refined. Therefore, this particular technique has not yet been used widely.

1.3 The Fédération des Établissements Hospitaliers et d'Aide à la Personne Privés Non Lucratifs (FEHAP) study from 2014

The FEHAP study is an antecedent study was carried out by colleagues in the same department as the author, in 2014, based on the same French hospital data from 2012 (9,15). It investigated unidirectional patient transfers from MCOs to SSRs in three of the former regions of France: Bretagne, Lorraine and Rhône-Alpes. It made use of various statistical techniques, including unweighted ERGM analyses. The main findings were that it confirmed the existence of regional differences in hospital transfer networks. The study also found that in the region of Rhône-Alpes, the links within the network depended the most on the sector of the sending and receiving hospitals being the same. However, in Lorraine and Bretagne, the main determinants of the shape of the network were travel times between the sending and receiving hospitals, and whether the two hospitals belong to the same geographical department. At the time of this original study, there was an unfulfilled wish to take the weight of transfer trajectories into account when examining the hospital transfer network. Namely, what determines how many patients are sent from one hospital to another? However, as discussed above, the techniques for performing this type of analysis were still nascent, and troubled by issues such as non-convergence, or in other words, failure to generate a result after protracted computation time. With further software developments, this may now be a more appropriate time to evaluate the weighted ERGM technique with these data.

Clearly, the best method for testing and understanding a new statistical technique is to apply it to real data. This current project is well-placed to do so, as it takes advantage of the previous study undertaken by the same academic department. This means that the same data will be re-examined with a novel, but highly related technique of weighted ERGM, with the results from the previous analysis acting as an informative comparator. This allows a clear sense of the contribution made by this new analysis.

1.4 Objectives

This project, to be carried out as part of a Master of Public Health dissertation, aims to offer a small but useful contribution to the evolving field of ERGM. Namely, its main objective is:

> To examine the usefulness of weighted ERGM when applied to inter-hospital patient transfers in the French health system.
>
> This entails the following component objectives:
>
> - To explore the determinants of acute-to-subacute patient transfers in three different regions of France, paying attention to the volume of patients in each transfer pathway
> - To assess regional differences in patient transfer patterns, again taking the volume of transfers into account
> - To compare analysis with the weighted ERGM technique against the original unweighted ERGM analysis in terms of additional information that could be provided, or the degree of fit to observed data

As with any dynamic national institution, there have been annual reforms and adjustments to the French healthcare system, ranging from focal changes to national re-structuring (11). As well as this, due to national territorial reforms in 2016, two of the three original regions, Lorraine and Rhône-Alpes, no longer exist with their 2012 borders. As such, this project, with hospital data dating from 2012, is designed for the purpose of evaluating new methods of studying inter-hospital patient transfers in France, rather than for advising stakeholders on planning and policy. Nevertheless, it is essential for studies like this to be undertaken, in order to assess the applicability of new methods before they are adopted widely. Potential future developments could include using weighted ERGM on more contemporaneous data, in order to influence policy and planning more directly.

2. Materials and Methods

2.1 Ethical Review

The ethical approval for this project was duly obtained from the Research Ethics Committee at the University of Sheffield School of Health and Related Research. As per the ethics protocol of the University, only a Secondary Data Declaration was required, as the project makes use of anonymised secondary data. (Appendix 10).

2.2 Data Sources

2.2.1 Patient-level data

The data used for this project were obtained from several sources. At the patient level, already-anonymised data were obtained from the *Programme de Médicalisation des Systèmes d'Information* (PMSI) database for 2012. The appropriate approval to use these data was granted by the *Commission Nationale de l'Informatique et des Libertés* (Appendix 10). The process of extracting information from this database was undertaken at the time of the original study in 2014. This extracted data contains information on 279,071 unique hospital stays across all of France in the year 2012. A list of the patient-level variables that were originally extracted can be found in Appendix 2.

Patients were then selected with the project inclusion criteria, as follows:

- Adult patients aged 18 or above
- Transferred from an episode of MCO care to an episode of SSR care. This project follows from the FEHAP study in focusing on transfers in this direction only, because transfer numbers from SSRs to MCOs are much smaller, and may be for a broad range of reasons that cannot be captured in the current dataset (9).
- Both the MCO and SSR stays taking place within the regions of Bretagne, Lorraine, or Rhône-Alpes
- Speciality of care within the major categories of neurological or musculoskeletal illness. This is also one of the criteria from the FEHAP study, and this particular decision was made because patients within these specialities represented the largest group of transfers from MCO to SSR care (9,15). This project has followed this in order for results due to a similar rationale, and also in order for results to be comparable against the FEHAP study.

2.2.2 Hospital-level data

Data on the particulars of each hospital were obtained from the *Statistique Annuelle des Établissements de santé* (SAE; Annual Statistics of Health Establishments), a publicly available and searchable database (21). Data for public, private and ESPIC hospitals and clinics in all regions of France in the year 2012 were retrieved. A number of data dictionaries were consulted, both within the SAE itself, and on the *Fichier National des Établissements Sanitaires et Sociaux* (FINESS;

National Catalogue of Health and Social Establishments), which is also available to the public (22). This process occurred during the construction of the dataset for this particular project, in 2019. A list of the variables chosen and retrieved can be found in Appendix 2.

SAE information was self-reported by each hospital, with little enforcement regarding the accuracy or completeness of each entry (9). Where there were suspected anomalies or omissions in hospital-level data, this was cross-referenced against FINESS. In the case of a small number of hospitals, the only option was to resort to consulting information supplied by the websites of the establishments themselves or to publicly available government documents published online. As such, there may be inaccuracies due to changes in hospital characteristics since 2012.

2.2.3 Geographical data

Consideration was given to using the original travel time data from the FEHAP study, as bias may be introduced if current travel times were applied to geographical data from 2014. However, as this project required travel times for all transfer trajectories, and not only those above a threshold weight, it was more practical to recompute all of the travel times together.

New travel time data were retrieved from *OpenStreetMap,* a publicly available source, which enabled the computation of travel times between the centroids of the communes where each hospital was located (23). As with other hospital-level data discussed above, the commune information for each hospital was obtained from SAE and FINESS. However, in the case of Lorraine, there is a cluster of ESPIC hospitals under the same umbrella organisation that share a single administrative postcode, despite being physically located in a number of departments in the region. Also, in the case of Rhône-Alpes, all of the hospitals belonging to the *Hospices Civils de Lyon* (the public hospital organisation of Lyon) are grouped geographically as a single entity, with a single administrative postcode identifier, which would create inaccurate travel time information. For these clusters, there was sufficient information available to create centroids, weighted by the number of beds at each individual establishment within the cluster.

The process of obtaining travel times was automated for each region by making use of the `osrm` package in R (24). These travel times were initially computed in minutes. A small sample of these new travel times were compared against those from the 2014 analysis, and any discrepancies were found to be within 5 minutes.

2.3 General study design

The outcome of this project is the observed structure of regional hospital transfer networks themselves, with the predictors being the patient-level and hospital-level variables described above. As described, the travel time data was collected in 2019, and all other variable data are over the year 2012. As such, the study design can best be described as a cross-sectional analysis of secondary data.

2.4 Adjacency matrices and network building

Data are typically read and converted into networks through several formats. One of the most common formats is an adjacency matrix, otherwise known as a transition matrix, where usually, the sender nodes are represented as rows, and the receiver nodes are represented as columns (25). In this project, adjacency matrices were created for each of the three regions, based on patient-level information. The sum of patients travelling along each unique edge, from one particular hospital to another, was incorporated in each matrix as the weight. From each of these adjacency matrices, two networks were built: one containing loops, and one without loops. As stated previously, loops in this context denote self-transfers of patients within the same hospital. In the network without loops, the weight of the self-transfers, which appear in each adjacency matrix as the top-left to bottom-right diagonal, is set at zero. All other features of the networks remain the same between the two versions. The decision to examine the networks both with and without loops is described in the Discussion section.

For each of the networks, with and without loops for each region, weighted assortativity was examined by both legal status and node degree (Section 2.5). Models were then built for each network (Section 3.3), and analysed using both unweighted and weighted ERGM techniques.

2.5 Assortativity analysis

Assortativity is a measure of the similarity of the two nodes at either end of an edge, in terms of a particular attribute (26). It is usually quantified, as a summary for the entire network, by the assortativity coefficient, which takes a value between -1 and +1. This coefficient is essentially a comparison of the observed network against a random network. Therefore, a positive assortativity coefficient denotes that compared to a random network, the nodes of the observed network are more likely to form edges between each other if they are the same in terms of a particular attribute. By the same logic, a negative assortativity coefficient means that nodes that are different in terms of a particular coefficient are more likely to form edges between each other, compared to those in a random network. An example where negative assortativity may occur can be partner selection, with relation to the attribute of gender. We would expect that in general, individuals of different genders may be more likely to select each other as partners, as compared to random pair selection.

In the analysis for this project, assortativity is calculated for the attributes of legal status and node degree. Assortativity by legal status reveals whether there is a greater tendency for patient transfers between hospitals that have the same legal status. Assortativity by degree may be more abstract. Essentially, if degree is the total number of edges entering and leaving a node, then it can be seen as a measure of the popularity of the node. Therefore, the question being posed by assortativity by degree is whether less popular hospitals are more likely to form patient transfer relationships with more popular hospitals, and vice versa.

In the original FEHAP study, assortativity by legal status and by degree were both analysed for edges at or above a threshold weight of 3 only (9,15). In the early stages of data exploration for this project, these results were reproduced. However, this analysis was then extended to include all edges and to make use of the weighted method for calculating assortativity, where the question became whether nodes that are more similar in the attributes of legal status and degree tended to have stronger relationships, in terms of higher numbers of patient transfers, than nodes that are less similar in these attributes. The weight of the patient transfer relationships between hospitals is a theme that is examined through a number of angles in this project.

The method for calculating assortativity by weight has only recently been developed, and is an extension of the conventionally accepted method for calculating assortativity for binary ties (27). The strength of the assortativity coefficient has been conventionally interpreted in the same manner as the Pearson’s correlation coefficient, where an absolute value of 0 indicates no association, and an absolute value of 1 indicates perfect correlation (25). In order to determine the significance of the assortativity values, standard errors were calculated using the jackknife method, as is becoming increasingly common in the network literature (26,27). In the project, the assortativity coefficient was considered to be statistically significant if the null value of 0 was excluded from 2 standard errors of the coefficient estimate.

2.6 ERGM analysis

As previously stated, ERGM is a technique for analysing network data. Various extensions to the basic ERGM techniques have become available in recent years. Coefficients given by the ERGM model summary can be interpreted as the log odds of there being an edge in the network. In other words, how much more likely an edge would form between two vertices, given a unit increase of a particular predictor (19).

ERGM analyses involve comparing a particular network that is observed in real life against the set of alternative networks that may be possible (25). It makes use of Markov Chain Monte Carlo (MCMC) methods, which are stochastic in nature. The Markov chain is an extensive sequence of random, stepwise changes, known as proposals, in the underlying adjacency matrix (28). This generates a large number of possible networks based on the number of nodes in the matrix, and any additional constraints specified. In the possible sample space of an unweighted network, edges are binary, which is to say that they can either exist or not. In a weighted network, edges can take numerical values. However, with the current development of the algorithm for analysing weighted ERGM models, there is an assumption of “infinite sample space” for edge weights (29). This means that the user cannot specify a more realistic upper bound on the edge weights, and therefore the algorithm cannot reject the possibility that there may be an edge with a weight of infinity. Due to this unboundedness, weighted models are far more computationally intensive than unweighted ones, as will be explored in greater detail in the Discussion section. This also necessitates adjustments to many of the control settings when the model analysis is performed.

2.6.1 MCMC diagnostics for ERGM

Even though the model output may appear to report convergence during the running process, this simply indicates the fulfilment of one of the key aspects of the Monte Carlo Maximum Likelihood Estimation (MLE) procedure. This does not equate to the convergence of the entire MCMC ERGM model, which also requires assessment through post-hoc statistics and through plots (30). Several statistics are generated after running each model, for the final iteration of simulations, immediately before the final parameter estimates are calculated. Among these, the autocorrelation statistics are computed. Autocorrelation is a measure of the similarity of a particular Markov chain sample to a previous sample drawn by the same chain. It is likely to be higher with a larger dataset, and a Poisson distribution, as compared to a Normal or a Binary distribution (31). This number should be minimised, as random sampling is desired. However, there are no rules regarding the maximum acceptable autocorrelation. Geweke's statistics are also given. These assess whether the length of the burn-in process is adequate, which assists in improving the likelihood of convergence of the model, but is not diagnostic of convergence in itself (30,32).

In terms of MCMC plots, it is important to assess the trace of the Markov chain visually, to ensure that it is homogenous throughout the duration of the chain and that it has covered a consistent region of the sample space (30). The density plot is also given in order to check for a unimodal and even distribution of the estimates obtained from sampling.

2.6.2 Goodness of Fit for ERGM

Goodness of fit (GOF) can be assessed for unweighted networks only (20). In the project, this was performed by an automated process simulating 100 networks using the parameters obtained through the MLE, and determining how well these 100 simulated networks match the observed network (25,28). The end statistics can be assessed visually by requesting GOF plots. GOF plots are typically created for a number of network statistics, such as in-degree and out-degree (term definitions in Appendix 1). Plots are determined to be better if the solid plotted line of the observed network falls within the dotted 95% estimate boundaries of the simulated networks (Appendix 8).

2.6.3 Unweighted ERGM

The original unweighted ERGM from the FEHAP study were repeated using the `ergm` package in R, but with the notable difference of including all edges, rather than only those that have sustained 3 or more transfers (9,15,28). The main reasons for making this decision are twofold. Firstly, for the hypothesis being considered in this particular study, where all eligible patient transfers are being examined, any edges along which a patient was transferred would be considered to be important for analysis. Secondly, in order to make meaningful comparisons between unweighted and weighted ERGM models, the underlying network for both analyses should be as similar as possible. Removing edges with weights less than 3 would not only reduce the number of edges in the underlying network, but it would induce nodes that are only connected by lighter edges to become isolates, thus changing some of the other statistical properties of the network.

For this project, in the unweighted ERGM algorithm itself, the maximum number of MCMC iterations was set empirically at 50, and this was found to be sufficient for the regional unweighted network models to reach convergence for the Monte Carlo MLE component. Within each iteration, the burn-in period was set at 100,000. This means that the programme is asked to run the Markov chain through 100,000 stepwise proposals initially, but these results are discarded, and not used in calculations, in order to allow the Markov chain to wander into a region of the sample space where it is sufficiently stable (30). After this period, random samples begin to be taken and used to calculate summary statistics. For this project, the sample size for the unweighted models is set at 100,000, and the sampling interval remains at the programme default of 1024. This means that a random sample is taken every 1024 proposals of the Markov chain, and a total of 100,000 random samples are taken. Between each iteration, the parameters within the algorithm change based on the results of the previous iteration. After convergence of the Monte Carlo MLE is achieved, the ergm algorithm ends by evaluating the marginal likelihood with a bridge sampling procedure. When the results of each model run are reviewed, a series of post-hoc tests are also performed to determine model convergence and to assess GOF, as described previously.

2.6.4 Weighted ERGM

As this is predominantly a project to examine weighted ERGMs from a standpoint of methodological adequacy for analysing patient transfers, appropriate comparisons between unweighted and weighted ERGMs needed to be made. Therefore, each of the unweighted ERGM models were repeated with weighted ERGM techniques. The weighted ERGMs were performed with the ergm.count extension of the ergm package in R (29). The weight of each edge corresponds directly to the total number of patients transferred between the two hospitals at either end of the edge during the year 2012. All other features of each network remain the same between the unweighted and weighted analyses.

The ergm.count extension allows the consideration of edge weight by comparing the observed graph to a random graph whose edges can take weights according to a binary, geometric or Poisson distribution (20). The Poisson distribution was used as reference for the models in this project, as this was the most appropriate choice for the hypothesis and data type. The ergm.count algorithm begins with up to 60 iterations of Contrastive Divergence Monte Carlo MLE (CD-MCMLE), which is for the purpose of determining a suitable starting value for the Markov chain in the next phase (33). Convergence is not always attained after this process. After this step, MCMC MLE iterations are run according to the same principle as for the unweighted ERGM. Starting parameters for unweighted ERGM analyses are usually determined empirically. In particular for this project, the burn in period was increased to 200,000, meaning that the first 200,000 proposals of each Markov chain are discarded before sampling is performed. The maximum number of MCMC MLE iterations was also increased, initially to 65, in order to allow an additional buffer in terms of sufficient runtime for convergence. This was later found to be inadequate for some of the larger networks in the project, and had to be adjusted upwards to 75. The sample size was set at 200,000, and the sampling interval was increased to 10,000 in order to reduce autocorrelation (34). This means that a total of

200,000 samples were taken at intervals of 10,000 proposals (see Appendix 6 for sample code). The final step of the ergm.count algorithm involves evaluating the marginal likelihood with bridge sampling, as occurs for the unweighted ERGM algorithm.

In terms of post-hoc testing, MCMC diagnostics can still be requested for a weighted ERGM, checking the same parameters as those performed for an unweighted ERGM. However, there are no goodness of fit statistics, as this is still poorly defined for weighted ERGMs (20).

2.7 Software and technology

The data management and modelling within this project were carried out using R software, versions 3.5.1, 3.5.2 and 3.6.0, within the RStudio interface. A list of packages and versions used can be found in the Appendix 3.

3. Results

3.1 Descriptive statistics

In total, 54,889 transfers were analysed across 2,663 unique edges, in the regions of Bretagne, Lorraine and Rhône-Alpes (Table 2).

Table 2: Descriptive statistics of regional networks with loops

	Bretagne	Lorraine	Rhône-Alpes
Number of hospitals (nodes)	89	80	199
- Number of public hospitals (%)	43 (48.31)	38 (47.50)	95 (47.74)
- Number of ESPIC hospitals (%)	24 (26.97)	27 (33.75)	47 (23.62)
- Number of private hospitals (%)	23 (24.72)	15 (18.75)	57 (28.64)
Total number of hospital beds	14,287	11,269	28,172
- Number of MCO beds (%)	9,979 (69.85)	8,237 (73.09)	19,036 (67.57)
- Number of SSR beds (%)	4,308 (30.15)	3,032 (26.91)	9,136 (32.43)
Number of edges	556	513	1,594
Number of patient transfers	16,548	9,260	29,081
Median patient transfers per edge	5	3	3
Mean degree (standard deviation)	12.49 (8.05)	12.83 (8.76)	16.02 (12.80)
Density	0.07	0.08	0.04

In terms of hospitals (89) and total bed numbers (14,287), Bretagne is marginally larger than Lorraine, and much smaller than Rhône-Alpes. However, it has a much larger number of patient transfers (16,548) than one may expect, given that the number of distinct edges (556) is only slightly higher than that for Lorraine. This means that the number of patients transferred along each edge would be higher, and indeed this is the case, with a median edge weight of 5, as opposed to 3 for the other regions.

At 80 hospitals with a total of 11,269 beds, and with 9,260 patient transfers along 513 edges, Lorraine is the smallest of the 3 regions. However, it has a particularly high percentage of ESPIC hospitals (33.75%) and a low percentage of private hospitals (18.75%). In terms of the distribution of beds, it has a relatively low percentage of SSR beds (26.91%). Of the 3 regions, it also has the highest network density.

As expected, Rhône-Alpes, as the most populous region, also has the most hospitals and patient transfers. Compared to Lorraine, it has more than double the number of MCO beds (19,036), and more than triple the SSR beds (9,136). This generated more than 3 times the number of patient transfers (29,081) in 2012. However, the median number of patients transferred along each edge during the year is the same as that for Lorraine, at 3, meaning that the number of edges is much greater, at 1,594, also more than 3 times that of Lorraine, and 2.9 times that of Bretagne.

In networks without loops, the number of hospitals and beds do not change. However, the number of edges has decreased to 2,514, as each loop represents both an in-edge and an out-edge for a hospital. The overall number of patients transferred has also decreased to 36,900, which is 67.23% of all of the networks with loops. Therefore, there are also changes to the network statistics (Table 3).

Table 3: Descriptive statistics of regional networks without loops

	Bretagne	Lorraine	Rhône-Alpes
Number of hospitals (nodes, unchanged)	89	80	199
Number of edges	513	477	1,524
Number of patient transfers	10,468	5,983	20,469
Median patient transfers per edge	4	2	3
Mean degree (standard deviation)	11.53 (7.82)	11.93 (8.56)	15.32 (12.67)
Density	0.07	0.08	0.04

With the removal of loops, certain network characteristics become more pronounced. For example, in Bretagne, the number of patient transfers is substantially higher (10,468) than that for Lorraine, However, compared with the looped version of itself, this represents a 36.7% reduction in transfers, suggesting that a considerable proportion of patient transfers in Bretagne occurs through hospitals sending patients to themselves. Despite this reduction, Bretagne still has the highest median patient transfers per edge, at 4.

Lorraine remains as the smallest region in terms of its non-loop edges (477). Similar to Bretagne, there is a 35.4% reduction in the number of patient transfers (5,983) once loops are taken out of consideration. This leads to the median patient transfers per edge decreasing to only 2.

Rhône-Alpes still remains by far the largest region by edges (1,524) and transfers (20,469). This is size difference is further amplified by the fact that the region has only experienced a 29.6% reduction in patient transfers, which means that its median patient transfers per edge remains at 3.

3.2 Assortativity

As described in section 2.5, assortativity means that nodes that are more alike in a particular attribute are more likely to interact with each other. In the networks with loops, weighted assortativity by node degree was only statistically significant in a positive direction for Bretagne. However, this was of a negligible magnitude. In contrast to this, weighted assortativity for the attribute of legal status was found to be statistically significant, also in a positive direction for the weighted networks of all regions (Table 4). Additionally, for a categorical variable such as legal status, the weighted assortativity algorithm generates an additional mixing matrix, which shows the assortativity coefficient for each possible pairwise combination of hospitals (Appendix 5). It can be seen that in the networks with loops, the strongest assortativity values are between public hospitals and other public hospitals.

In the networks without loops, there was only statistically significant positive assortativity for node degree in Rhône-Alpes (Table 4). This result is not statistically significant in the corresponding network with loops. There is no assortativity for legal status in any of the regional networks without loops. Furthermore, the mixing matrix does not indicate any particular patterns nor outliers (Appendix 5).

Table 4: Assortativity of regional networks

	Bretagne		Lorraine		Rhône-Alpes	
	With loops	No loops	With loops	No loops	With loops	No loops
Assortativity of node degree (SE)	0.38* (0.14)	-0.05 (0.07)	0.26 (0.15)	-0.09(0.08)	0.55 (0.32)	0.14* (0.06)
Assortativity of legal status (SE)	0.25* (0.06)	0.05 (0.05)	0.31*(0.07)	0.06(0.05)	0.28* (0.05)	0.08 (0.04)

** Denotes statistical significance at 2 standard errors*

3.3 Variable selection and ERGM model building

3.3.1 Selection of additional potential determinants of patient transfer

After selecting only those patients who fit the selection criteria, additional hospital-level aggregates were created from some of the patient-level data, as it was thought that they represented some of the other potential determinants of inter-hospital patient transfer:

- Mean age of patients
- Median age of patients

Univariate statistics were performed for these variables (Table 5), and it was found that the mean and median of both of these variables were well above the retirement age cutoff of 65, which would have been a clinically meaningful value. The sample size below age 65 was too small to allow a reasonable analysis. Therefore, a decision was made not to use either of these variables in the model analysis.

- Proportion of female patients

A decision was made to create this variable as a gender proportion instead of a ratio, given that some hospitals may have very low numbers in 2012, where any gender imbalances may create very extreme values in a sex ratio. Univariate statistics were performed for this variable, and it was found that most hospitals had a very high proportion of female patients. Therefore, it was not meaningful to compare hospitals with respect to their patient gender proportions, and again, a decision was made not to use this variable in the model analysis.

- Mean LOS for MCO care
- Median LOS for MCO care

Once again, univariate statistics were also performed for these variables. For the model analysis, the *median MCO LOS* variable was chosen over the *mean MCO LOS*, because its distribution was slightly less skewed (Table 5).

Table 5: Descriptive statistics for additional variables

	Bretagne	Lorraine	Rhône-Alpes
Mean age of patients (hospital aggregate)*			
- Median value of hospital mean age (Range)	77.22 (31.55- 84.03)	77.81 (45.50-84.82)	76.45 (19.00-86.99)
- Number of hospitals mean age ≥65 (%)	81 (94)	73 (95)	178 (92)
- Number of hospitals mean age <65 (%)	5 (6)	4 (5)	16 (8)
Median age of patients (hospital aggregate)*			
- Median value of hospital median age (Range)	80.00 (31.00-85.00)	80.15 (47.00-85.91)	79.74 (19.00-88.00)
- Number of hospitals median age ≥65	82 (95)	73 (95)	182 (94)
- Number of hospitals median age <65	4 (5)	4 (5)	12 (6)
Proportion of female patients*			
- Number of hospitals with proportion female > 50 (%)	80 (93)	75 (97)	185 (95)
- Number of hospitals with proportion female ≤ 50 (%)	6 (7)	2 (3)	9 (5)
Mean LOS for MCO care			
- Median value of Mean MCO LOS (Range)	12.43 (4.25-36.20)	12.15 (6.88-26.25)	11.72 (5.73-31.33)
Median LOS for MCO care			
- Median value of Median MCO LOS (Range)	10.00 (2.00-27.00)	11.00 (6.00-25.50)	10.00 (5.00-31.00)

**3 unknown hospital values in each of Bretagne and Lorraine, 5 unknown hospital values in Rhône-Alpes*

It was decided to perform a formal diagnostic test of collinearity between the *median MCO LOS* and *MCO beds* variable, given that it could be reasonably argued that stay duration and bed availability may influence the time point at which an MCO may transfer a patient. The variance inflation factor (VIF), commonly used for checking collinearity, cannot be used on ERGM results, as ERGMs by definition violate the fundamental assumption of independence of variables required for the generation of VIF (35). Although various methods have been proposed to check for collinearity in ERGMs, none had been made widely available as an R package, especially for a weighted and directed network. Therefore, it was decided that for this project, for a general indication of potential collinearity, linear regressions would be performed in order to determine the VIF when *MCO beds* and *median MCO LOS* are jointly used to predict the mean network degree of each of the regional networks with loops.

Table 6: VIF values for network degree, with MCO beds and median MCO LOS as covariates

	Bretagne	Lorraine	Rhône-Alpes
VIF	1.04	1.01	1.01

As the VIF values are low for all regions, it is a reassuring indication that the choice of including *median MCO LOS* in ERGM models is less likely to be affected by any collinearity between this variable and the *MCO beds* variable.

3.3.2 Travel time matrix building

The travel times between hospitals were initially organised into one square transition matrix for each region, with rows representing the origin hospitals, and columns representing destination hospitals. This is a requirement of the ergm statistical package, where the status of every possible edge between every possible pair of nodes needs to be specified.

However, when this was applied to the analysis, it was found that the unweighted ergm algorithm will not run with travel time values of 0, and the remaining values were too heterogenous to be processed consistently. Therefore, after calculating the distribution of travel times (Table 7), a decision was made to discretise this variable according to quartiles. After this, the middle value of each quartile was used to represent the entire category, in order to comply with the requirement for the variable type to be numeric. This value was 10, 30, 50 and 70 for the four categories. An additional complication was that when loops were removed from the network, in order to allow the comparison of networks with and without loops, the lack of edge weights along the diagonals of the patient transfer matrices became exactly collinear with the lowest travel time category along the diagonals of the travel time matrices. This also impeded the running of the algorithm. To remedy this, a random number sequence was generated for each region, following a normal distribution with a mean of 5 and a standard deviation of 2, and rounded to the nearest positive integer. This sequence was then used to replace the diagonals of the travel time matrices. This process is described further in in the Discussion section, but essentially reflects the hypothesis that even within the same hospital, it may take on average 5 minutes, and probably up to 10 minutes, for a patient to be moved between wards. This procedure reduced collinearity in the matrices without loops to a point where the ergm algorithm could compromise by calculating maximum pseudolikelihood estimates (MPLE) for coefficients, which are less ideal than the intended maximum likelihood estimates (MLE) (36). It also allowed more reliable estimates to be generated for the networks with loops. This is because even without collinearity between travel times and transfer weights, travel times on their own are consistently zero along the diagonals of the matrices, and hence completely predictable. Therefore, the travel time matrices with random number diagonals were used for all analyses involving travel times in the project.

Table 7: Initial distributions of travel times in minutes, for networks with loops

	Bretagne	Lorraine	Rhône-Alpes
Median travel time	41.60	34.10	41.00
Range travel time	0.00 to 198.70	0.00 to 139.50	0.00 to 218.90

3.3.3 Model building and optimisation

Initially, models were constructed for the looped networks in accordance with those chosen in the 2014 FEHAP Study. This facilitates comparison between analysis with the original dataset and constraints, and the current one.

For the Bretagne region, all of the 4 original FEHAP models were built (Appendix 4)(9). However, it was decided that the models without class constraints, models 1 and 3, did not add useful information for the purposes of this project. Therefore, they were not constructed for the regions of Lorraine and Rhône-Alpes. Setting class constraints essentially means informing the algorithm that certain combinations of transfers are not possible, and therefore when random graphs are being generated, they should not contain any edges with forbidden combinations. In the case of this project, these combinations are edges travelling from an MCO to an MCO, from an MCO/SSR to an MCO, from an SSR to an MCO, and from an SSR to an MCO/SSR. The original models 2 and 4 from the FEHAP study specify class constraints, and have been adopted within this project, as model A "legal status" and B "travel times" (Table 8). Letters were used to name models in this project, in order to differentiate them from the FEHAP models, which are numbered. Another important difference to note is the omission of the variable *isolates*. This is due to the fact that the original study only analysed transfers with a weight of 3 or greater. Thus after some of the edges of weight 1 or 2 are removed, some nodes are no longer connected to any edges, becoming isolates. However, this project analyses every edge, and isolates are not created.

As well as this, additional models were considered in order to extend the original analysis, and to account for the idiosyncrasies of the computer algorithms. In particular, dummy bed variables, *no MCO beds* and *no SSR beds* were built for each network. This was because neither the ergm nor the ergm.count algorithms could process NA values for hospital beds, despite the fact that an MCO-only hospital would legitimately have NA values for SSR beds, and vice-versa. The dummy variables would be able to help examine the question of whether hospitals without MCO beds behave differently from the hospitals that do, and likewise with hospitals without SSR beds. This cannot be done simply by hospital type because there are 3 types of hospitals in the dataset: MCOs, MCO/SSRs and SSRs. These dummy variables were initially tested on the unweighted networks in the Bretagne region only, to determine if their addition produced statistically significant changes. Once this was confirmed, it was decided to construct model C "dummy beds" for all networks, with these additional dummy hospital bed variables, in addition to all of the other variables and class constraints already present in model B "travel times".

As already mentioned, the *median MCO LOS* variable was constructed and chosen specifically to test the hypothesis that the length of stay in an MCO facility would be a determinant for patient transfer. This was then added to all elements of model C "dummy beds" to create the final model, model D "length of stay".

Table 8: Final models constructed and analysed, per region

Model name	A: Legal status				B: Travel times				C: Dummy beds				D: Length of stay			
Networks	Loops		No loops		Loops		No loops		Loops		No loops		Loops		No loops	
Technique	U	W	U	W	U	W	U	W	U	W	U	W	U	W	U	W
Variables included:																
Legal status	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Travel time					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Department					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MCO beds					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SSR beds					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
No MCO beds									✓	✓	✓	✓	✓	✓	✓	✓
No SSR beds									✓	✓	✓	✓	✓	✓	✓	✓
Median MCO LOS													✓	✓	✓	✓

Key: U = Unweighted ERGM, W = Weighted ERGM; Shaded in grey = not included in model

3.4 Unweighted ERGM

Eight models of unweighted ERGM were ultimately performed for each region, as outlined in the previous section (Table 8 above). For many of the networks without loops, the MCMC MLE procedure for calculating standard errors did not converge. In these cases, the coefficient estimates are still available, and the ergm algorithm is able to provide the MPLE standard errors. This means that the coefficient estimates can still be used. However, MPLE standard errors are usually much wider than MLE standard errors, and considered to be unreliable (36).

The term *edges* is required by the unweighted ERGM model as the intercept term, and the estimate can be interpreted as the log odds of there being an edge between two nodes in this model, if the values of all of the other covariates are set as null or at their base values (34). For all models, the coefficient of the *edges* term is negative. This means that in all of these models, the observed network is less dense than would be expected in a completely random network with the same nodes.

The numerical variables in the models are *travel times*, *MCO beds*, *SSR beds* and *median MCO LOS*. Their coefficients can be interpreted as the log odds of there being an edge between two nodes in this model, if the variable is increased by one unit. For *travel times*, even though artificial categories were introduced, as described above, the units are still in minutes.

The categorical variables can be considered as two types of binary variables. The first type includes the *no MCO beds* and *no SSR beds* variables, which are coded as 1 if the hospital lacks a particular type of bed, and 0 if the hospital does not lack (and therefore has) a particular type of bed. The second type are the *legal status* and *department* variables. Although there are several categories within each variable, they are analysed using the "nodematch" command in ergm. This means that the value of the variable is assigned as 1 if nodes at either ends of an edge match in this variable, and assigned 0 if the nodes are discordant with regard to this particular attribute (28). The coefficient can be interpreted as the log odds of there being an edge between two nodes, if the value of the variable changes from 0 to 1. Therefore, as an example using Bretagne model D with loops, the log odds of there being a patient transfer between two hospitals of the same department is 3.97 times that of the log odds of there being a transfer between two hospitals of different departments.

It can be seen from the regional results tables that there is clearly a significant difference moving from models A to B, where a range of additional predictors are introduced. Moving from B to C, the *legal status* variable, in the Bretagne and Lorraine regions, is substantially affected by the addition of the *no MCO beds* and *no SSR beds* variables. However, this relationship is not seen in Rhône-Alpes, where it appears that *legal status* may have a more complex set of connections with a range of predictors. Models C and D tend to have very similar values for coefficient estimates. The difference between models C and D is the introduction of the median *MCO LOS variable*, which is only statistically significant for the network with loops for Rhône-Alpes.

Table 9: Unweighted ERGM results Bretagne

	A		B		C		D	
	Loops	No loops	Loops	No loops^	Loops	No loops^	Loops	No loops^
Edges	-1.96*** (0.06)	-1.96*** (0.06)	-9.10*** (0.47)	-18.80 (313.50)	-10.33*** (0.54)	-21.63 (308.50)	-10.98*** (0.78)	-22.20 (308.17)
Legal status	0.29** (0.09)	0.07 (0.09)	0.26 (0.18)	-0.84*** (0.25)	0.81*** (0.21)	0.01 (0.28)	0.80*** (0.21)	0.004 (0.28)
Travel time			0.30*** (0.02)	1.20 (31.35)	0.30*** (0.02)	1.23 (0.31)	0.30*** (0.02)	1.23 (30.82)
Department			3.89*** (0.29)	4.56*** (0.51)	3.96*** (0.29)	4.76*** (0.52)	3.97*** (0.29)	4.77*** (0.52)
MCO beds			0.001*** (0.0003)	0.001*** (0.0003)	0.001*** (0.0003)	0.002*** (0.0004)	0.001*** (0.0003)	0.002*** (0.0004)
SSR beds			0.0005 (0.002)	0.00007 (0.002)	0.004** (0.002)	0.006* (0.0004)	0.005* (0.002)	0.007** (0.0003)
No MCO beds#					0.36 (0.21)	1.23*** (0.27)	0.65* (0.32)	1.49*** (0.44)
No SSR beds##					1.29*** (0.23)	1.99*** (0.29)	1.41*** (0.26)	2.11*** (0.33)
Median MCO LOS							0.02 (0.02)	0.02 (0.03)

***Key:** Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05. ^ = Only MPLE SE available. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

Table 10: Unweighted ERGM results Lorraine

	A		B		C		D	
	Loops	No loops	Loops	No loops^	Loops	No loops^	Loops	No loops^
Edges	-1.80*** (0.06)	-1.80*** (0.06)	-9.10*** (0.51)	-19.26 (338.1)	-9.70*** (0.55)	-20.97 (334.80)	-9.82*** (0.66)	-20.79 (334.40)
Legal status	0.20* (0.10)	-0.01 (0.10)	-0.30 (0.19)	-0.58* (0.24)	0.48* (0.20)	-0.27 (0.25)	0.48* (0.20)	-0.27 (0.25)
Travel time			0.30*** (0.02)	1.20 (1.20)	0.30*** (0.02)	1.21 (33.48)	0.29*** (0.02)	1.21 (33.44)
Department			3.91*** (0.33)	5.01*** (0.72)	3.92*** (0.20)	5.03*** (0.72)	3.93*** (0.33)	5.03*** (0.72)
MCO beds			0.001*** (0.0003)	0.001*** (0.0003)	0.001*** (0.0003)	0.002*** (0.27)	0.001*** (0.0003)	0.002*** (0.0003)
SSR beds			0.0003 (0.002)	0.0003 (0.003)	0.003 (0.002)	0.004 (0.003)	0.003 (0.002)	0.004 (0.003)
No MCO beds#					0.21 (0.21)	1.16*** (0.0003)	0.25 (0.25)	1.09** (0.33)
No SSR beds##					0.74** (0.23)	1.21*** (0.28)	0.75** (0.24)	1.20*** (2.88)
Median MCO LOS							0.005 (0.01)	-0.008 (0.02)

***Key:** Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05; ^ = Only MPLE SE available. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

Table 11: Unweighted ERGM results Rhône-Alpes

	A		B		C		D	
	Loops	No loops	Loops	No loops^	Loops	No loops^	Loops	No loops^
Edges	-2.36*** (0.03)	-2.36*** (0.03)	-10.59*** (0.32)	-19.60 (185.20)	-10.61*** (0.33)	-20.97 (334.80)	-9.79*** (0.34)	-22.20 (308.17)
Legal status	-0.01 (0.05)	-0.13* (0.05)	0.40*** (0.11)	-0.15 (0.13)	0.39*** (0.11)	-0.27 (0.25)	0.55*** (0.11)	0.005 (0.28)
Travel time			0.40*** (0.02)	1.24 (18.52)	0.40*** (1.57)	1.21 (33.48)	0.40*** (0.02)	1.23 (30.82)
Department			4.66*** (0.19)	5.26*** (0.30)	4.67*** (0.19)	5.03*** (0.72)	4.67*** (0.19)	4.77*** (0.52)
MCO beds			0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)	0.002*** (0.0003)	0.001*** (0.0002)	0.002*** (0.0004)
SSR beds			-0.002 (0.0009)	-0.002 (0.001)	-0.001 (0.0009)	0.004 (0.003)	-0.003** (0.001)	0.007** (0.003)
No MCO beds#					0.08 (0.09)	1.16*** (0.27)	0.06 (0.09)	1.48*** (0.44)
No SSR beds##					-0.05 (1.1)	1.21*** (0.28)	-0.05 (0.11)	2.11*** (0.33)
Median MCO LOS							-0.05*** (0.009)	0.02 (0.03)

***Key:** Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05; ^ = Only MPLE SE available. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

Examining the networks without loops specifically, it can be seen that as expected, the MPLE estimates produced wide confidence estimates. This appears to affect edge attributes such as *edges* and *travel time* more than it affects the nodal attributes. In some of the nodal attributes, particularly *MCO beds* and *SSR beds*, there seems to be some similarities in the estimates produced by the networks with and without loops. In general, the *MCO beds* variable is more likely to be statistically significant compared to the *SSR beds* variable. This also applies to the *no SSR beds* variable, which can be considered as a counterpart to the *MCO beds* variable, in that many hospitals with MCO beds will not have SSR beds.

In terms of GOF, plots for all networks consistently showed a substantial improvement between models A and B, and a smaller improvement between models B and C. In Bretagne and Lorraine, there was no noticeable improvement between models C and D upon inspection of the GOF plots. In Rhône-Alpes, there may be a very slight improvement between models C and D (see Appendix 8 for an example of a GOF plot series).

3.5 Weighted ERGM

Eight models of weighted ERGM were performed for each region, corresponding with each unweighted model presented previously (Table 8 above). In the case of Rhône-Alpes, the main MCMC phase of the `ergm.count` algorithm did not converge within the specified number of iterations for the looped network for model B, and both networks for models C and D. In these cases, the algorithm gives the "best guess" estimates based on the final MCMC iteration performed (29). Therefore, the estimates themselves may not be accurate, as opposed to the case of the unweighted networks with no loops, where only the standard errors were unreliable. For this reason, most of the weighted ERGM interpretation in this project will be based on the Bretagne and Lorraine results.

In contrast to the unweighted ERGM model, the intercept term in a weighted ERGM is `sum` (29). Once again, in models C and D for all regions, this term is negative, suggesting networks that are less dense than would be expected if they were generated randomly.

With weighted ERGM, there are no GOF metrics to guide model selection. It can be seen from the coefficients themselves that there is a progression from models A to C, but many similarities between models C and D. Once again, going from model B to C, the *no MCO beds* and *no SSR beds* variables seem to have a notable effect on the *legal status* variable. With the introduction of the *median MCO LOS* variable in model D, it appears that the variable that is the most affected is *no MCO beds*.

Table 12: Weighted ERGM results Bretagne

	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Sum	1.00*** (0.01)	1.00*** (0.01)	1.92*** (0.035)	-1.97*** (0.035)	-3.07*** (0.04)	-3.41*** (0.04)	-3.40*** (0.08)	-4.01*** (0.09)
Legal status	0.40*** (0.02)	-0.38*** (0.02)	-2.99*** (1.99)	-0.53*** (0.02)	0.13*** (0.02)	-0.04* (0.02)	0.14*** (0.02)	-0.05* (0.02)
Travel time			0.02*** (0.0005)	0.02*** (0.0005)	0.01*** (0.0005)	0.01*** (0.0005)	0.01*** (0.0005)	0.01*** (0.0005)
Department			3.14*** (0.03)	3.05*** (0.03)	3.17*** (0.03)	3.07*** (0.03)	3.18*** (0.03)	3.08*** (0.03)
MCO beds			0.0009*** (0.0003)	0.001*** (0.0003)	0.001*** (0.00003)	0.002*** (0.00003)	0.001*** (0.0003)	0.002*** (0.0003)
SSR beds			0.004*** (0.0002)	0.004*** (0.0002)	0.005*** (0.0002)	0.005*** (0.0002)	0.005*** (0.0002)	0.005*** (0.0002)
No MCO beds#					1.02*** (0.02)	1.26*** (0.02)	1.17*** (0.04)	1.54*** (0.04)
No SSR beds##					0.90*** (0.02)	1.09*** (0.03)	0.94*** (0.03)	1.20*** (0.03)
Median MCO LOS							0.01*** (0.003)	0.02*** (0.003)

*Key: Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

Table 13: Weighted ERGM results Lorraine

	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Sum	-0.59*** (0.02)	-0.59*** (0.02)	-2.17*** (0.04)	-2.29*** (0.05)	-2.93*** (0.05)	-3.46*** (0.06)	-2.38*** (0.07)	-2.82*** (0.08)
Legal status	0.70*** (0.02)	-0.05 (0.03)	0.04 (0.04)	-0.27*** (0.03)	0.26*** (0.03)	0.07* (0.03)	0.27*** (0.03)	0.07* (0.03)
Travel time			0.03*** (0.0006)	0.03*** (0.0006)	0.02*** (0.0006)	0.03*** (0.0006)	0.02*** (0.0006)	0.03*** (0.0006)
Department			2.87*** (0.04)	2.76*** (0.04)	2.82*** (0.04)	2.70*** (0.04)	2.83*** (0.04)	2.71*** (3.83)
MCO beds			0.001*** (0.00003)	0.001*** (0.00003)	0.001*** (0.00003)	0.002*** (0.00003)	0.001*** (0.00003)	0.002*** (0.00003)
SSR beds			0.002*** (0.0002)	0.002*** (0.0002)	0.003*** (0.0003)	0.003*** (0.0003)	0.002*** (0.0003)	0.003*** (0.0003)
No MCO beds#					0.71*** (0.03)	1.07*** (0.03)	0.50*** (0.03)	0.80*** (0.04)
No SSR beds##					0.60*** (0.03)	0.86*** (0.03)	0.52*** (0.03)	0.79*** (0.03)
Median MCO LOS							-0.02*** (0.002)	-0.03*** (0.003)

*Key: Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

Table 14: Weighted ERGM results Rhône-Alpes

	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Sum	0.14*** (0.009)	0.14*** (0.009)	-2.24*** (0.0002)	-2.25*** (1.94)	-2.525*** (0.02)	-3.03*** (0.005)	-1.65*** (0.03)	-1.96*** (0.004)
Legal status	0.48*** (0.01)	-0.10*** (0.01)	-0.03* (0.01)	-0.11*** (0.01)	0.06*** (0.01)	0.47*** (0.01)	0.29*** (0.01)	0.50*** (0.01)
Travel time			0.03*** (0.0003)	0.03*** (0.0003)	0.04*** (0.0003)	0.04*** (0.0002)	0.04*** (0.0003)	0.04*** (0.0002)
Department			2.84*** (0.02)	2.91*** (1.71)	2.82*** (0.02)	2.11*** (0.01)	2.79*** (0.02)	2.09*** (0.01)
MCO beds			-0.0001*** (0.00002)	-0.11*** (0.01)	-0.00007*** (0.00002)	0.0007*** (0.00002)	0.0002*** (0.00002)	0.0007*** (0.00002)
SSR beds			0.004*** (0.0009)	0.004*** (0.00009)	0.003*** (0.00009)	0.002*** (0.00009)	0.002*** (0.00009)	0.002*** (0.00009)
No MCO beds#					0.25*** (0.01)	0.67*** (0.01)	0.10*** (0.01)	0.10*** (0.01)
No SSR beds##					0.24*** (0.01)	0.72*** (0.01)	0.002*** (0.00009)	0.61*** (0.01)
Median MCO LOS							-0.05*** (0.001)	-0.05*** (0.0009)

*Key: Estimates (Standard Errors), *** = p-value < 0.001; ** = p-value < 0.01; * = p-value < 0.05; Light grey shading = MCMC MLE did not converge and estimates are based on the final iteration of the algorithm. # = Reference group are hospitals with MCO beds; ## = Reference group are hospitals with SSR beds.*

3.6 Comparison between unweighted and weighted ERGM

As described above, the general patterns observed in the unweighted networks are generally repeated in the weighted networks. An additional similarity to note is that throughout all of the models, the *department* variable appears to be strongly influential. The *travel time* variable is very similar between equivalent networks of different regions. This is particularly noticeable in the unweighted networks.

The *legal status* variable diminishes in statistical significance going from each looped network to its corresponding non-loop network. However, given the uncertainty in determining statistical significance with the MPLE estimates for the unweighted networks without loops, the pattern is clearer in the weighted networks.

Interestingly, in contrast to the unweighted models, the *median MCO LOS* variable is statistically significant in the weighted models, but with positive coefficients in the Bretagne region, and negative coefficients in the Lorraine region.

4. Discussion

4.1 Methodological considerations

Several key decisions needed to be made in the process of preparing the data for use in ERGM. This was both in order to address specific study questions, and to adapt to the requirements of the algorithms themselves. This means that choices were necessary, and resulted in certain assumptions being introduced into the analysis.

4.1.1 Unipartite vs bipartite networks

One of the more crucial initial decisions in this project was whether to construct the networks as unipartite or bipartite. In the case of a unipartite network, all hospitals would be considered to belong to one general type. However, a decision could also have been made to consider the existence of two distinct types of hospitals: those offering MCO care and those offering SSR care. In this case, constraints could be made in that patients can only be transferred from an MCO to an SSR, and not between hospitals of the same type.

The decision to represent the networks as unipartite was mostly due to the methods available to investigate bipartite graphs. In order to perform weighted ERGM analyses on bipartite graphs, the use of a technique known as projection would have been necessary (17). Essentially, this would involve splitting each network into separate MCO-only and SSR-only networks, and in doing so, eliminating the actual patient transfer pathways that are the main focus of study in this project.

4.1.2 Networks with and without loops

The rationale for analysing models both with and without loops involves several considerations. Using loops in models permits the full dataset to be captured in the analysis. It allows the consideration that in reality, a large proportion of the patient transfers between MCO and SSR care in each network occur via loops. However, drawbacks are that out of all of the hospitals, those that are able to have loops, meaning those hospitals that are able to provide both MCO and SSR care, may be fundamentally different to those that are not. Furthermore, the factors that drive patient transfers within a hospital may be different from those that influence transfers between hospitals, and these subtleties may be lost if all hospital transfers are analysed as a whole. In terms of accounting for these potential issues, the choice of strategy is constrained by methodological limitations. For example, the ergm.count package is as yet unable to take into account binary categorical edge attributes such as whether an edge is a loop or is not, and therefore there cannot be an additional term relating to edge type in the model equation (29). For the same reason, there also cannot be a binary term for whether the travel time for the patient transfer is zero or not. Another strategy would be to create additional dummy hospitals for those that have self-loops, so that each transfer can be viewed as a journey between the sending half of the hospital, and the receiving half. However, this would entail fundamental transformations in the underlying adjacency matrices, which would be difficult to construct practically. Therefore, the most straightforward method to assess the impacts of loops on the networks is to construct networks without loops to match those with loops, and to examine the differences between them.

The networks without loops are useful for examining the determinants of inter-hospital patient transfers specifically, which would occur when acute care hospitals are unable to provide subacute care and need to transfer patients to other facilities. As well as this, it leads to more homogeneity in the patient transfers being analysed, and reduces confounding factors that are unable to be captured by the variables available in the dataset.

4.1.3 Adjustments to travel times

As already described above, much manipulation was required to create travel time matrices that could be used by the ergm and ergm.count algorithms. It was necessary to introduce artificial categories for a continuous variable. The decision to use the centre point for each time category means that the lowest category was represented by a travel time of 10 minutes. This was justified by the fact that in real life, this could be a reasonable transit time even within one hospital, between one ward and another.

Furthermore, the need to generate random numbers between 1 and 10 for the diagonal of the travel time matrix means that error was deliberately incorporated into the data. On the other hand, this can also be viewed as being representative of the random variation in transit time within hospitals.

However, despite the justifications for the rationale behind this treatment of the travel times variable, it can be seen that the end result could be artifactual from these manipulations, in terms of the similarities between regions. Therefore, an area for future development in ERGM analyses of patient transfer data would be to find an improved method for investigating travel times between hospitals, especially in networks without loops.

4.2 Comparisons between unweighted and weighted techniques

In general, the unweighted and weighted ERGM techniques generated concordant results. Therefore, in terms of choosing between these techniques, important considerations may be the size and complexity of the network, and whether the analysis is intended to be used for description or prediction.

4.2.1 Convergence

Model convergence is a key consideration with ERGM. At various points during this project, either the unweighted or the weighted ERGMs have failed to converge, often due to different reasons. The unweighted ERGM algorithm appears to be more sensitive to certain features of the data structure, such as the presence of NAs (19). It is also more easily affected by particular characteristics of edge attributes, such as collinearity between zero travel time and zero weight in the networks without loops. When the MCMC MLE process within the unweighted ERGM algorithm fails to converge, it will still return valid estimates, but with wider and less reliable MPLE standard errors (36).

In the case of weighted ERGM, convergence seems to be more related to the volume of information, including the number of nodes and edges in the underlying network, and the number of covariates in the model. A model with a smaller network and only a few covariates, such as Lorraine model A, converged within 18 MCMC iterations. Conversely, in the largest region, Rhône-Alpes several models did not converge within 65 or 75 iterations. This finding is not unexpected, in that the algorithm needs to compute a greater number of estimates over an unlimited sample space. When non-convergence occurs, the unweighted algorithm returns the estimates from the final MCMC iteration that was performed, which may not be accurate nor reliable.

4.2.2 Goodness of fit

As mentioned earlier, there is currently no way to assess the GOF of weighted ERGMs. This is a major disadvantage of using the weighted ERGM technique, in that this makes the comparison between models difficult. This would also limit the usefulness of weighted ERGM in simulations and projections. The GOF assessment for an unweighted network cannot be used as a proxy for the GOF of a corresponding weighted network, because weighted networks represent different models altogether. The spectral goodness of fit method has been suggested (37). However, it has not become widely accepted, and the statistical package to perform this is currently unable to process directed networks. As well as this, some statisticians regard the visual assessment of GOF plots for unweighted ERGM as being crude, and needing to be replaced by statistical tests (19,38).

4.2.3 Computation time

The main objective of this project was to explore the usefulness of the weighted ERGM technique for examining patient transfers. As such, a major practical consideration is the computation time required by each run of the algorithm. However, a considerable amount of flexibility is available through the control parameters of the algorithm, in terms of the number of random networks to be generated or random samples to be taken. Due to particular idiosyncrasies of the data used in this project, such as the directed and weighted properties of the edges, the constraints between different classes of hospitals, and the toggling of loops, more iterations may be required. A number of computers were used to run the weighted ERGM models in this project. The computation times are given for reference in Appendix 9.

4.3 Determinants of patient transfer

As mentioned in the Results section, each of the regional networks were found to be less dense than randomly generated networks with the same nodes, with both the unweighted and unweighted techniques. This is also corroborated by the initial network density statistics. It has been suggested that a less dense hospital network may be an indication of a better organised system, where there is an established hierarchical or preferential system dictating the flow of patients, so that sender hospitals do not simply transfer to all other possible receiver hospitals at random (5).

The initial weighted assortativity analysis suggested that the presence of loops may have an effect for some of the predictors. The positive assortativity for *legal status* observed for the networks with loops in all regions became statistically insignificant when loops were removed. This is explained by the fact that if hospitals are sending many patients to themselves, then clearly it follows that they are more likely to associate with hospitals that have the same legal status as themselves. However, when only the relationships with other hospitals are examined, it can bee seen that the legal status of other hospitals are not important in terms of whether edges are formed. Similar patterns can be seen in both the unweighted ERGM results. Focusing on unweighted models C and D for all regions, it can be seen that the *legal status* variable is only statistically significant for the networks with loops. In the weighted models C and D, this effect can also be seen, but to a lesser extent. Furthermore, with regard to *legal status*, it can be seen that there is some relationship between this variable, and the *no MCO beds* and *no SSR beds* variables. With the addition of the no beds variables between models B and C, there are dramatic changes in the magnitude, and sometimes the direction, of the coefficient estimates for *legal status*. In the original FEHAP study, the analysis finishes at the equivalent of model B of this project, where it was found that legal status was not significant for Bretagne and Lorraine (9,15). It can be seen that extending the analysis with weights, networks without loops and additional variables has led to different results, and has changed the interpretation of the role of this variable.

Another variable that attracted attention in the FEHAP study was *department* (9,15). It had been found to be one of the strongest determinants of patient transfer. This was also the case in this project, in both the unweighted and weighted networks. This may have some relationship with the residential address of the patients themselves, and may be an interesting direction for future research. Interestingly, in the unweighted networks, the strength of this variable increases from looped networks to non-loop networks, whereas in the weighted networks, the opposite occurs. Unweighted networks can be seen as reflecting the number of relationships between hospitals, and weighted networks as the strength of these relationships. Therefore, it follows that in the unweighted networks, when self-transfers are excluded, the remaining transfers tend to be to a large number of hospitals within the same department. However, in the weighted networks, it can be seen that when self-transfers are removed, the remaining number of patients being transferred to hospitals within the same department is much smaller.

In terms of the *MCO beds* and *SSR beds* variables, it can be seen that despite the very small coefficients, patient transfers tend to be more likely between larger hospitals than smaller ones. The coefficients became more positive going from networks with loops to those without loops. This suggests that when only inter-hospital transfers are examined, larger hospitals were even more likely to be involved.

A similar pattern can be seen in the *no MCO beds* and *no SSR beds* variables. Hospitals that are coded as 1 for either of these variables are unable to transfer patients to themselves, as they lack one bed type. However, when loops are removed, then as expected, these hospitals become much more likely to participate in patient transfers.

As previously stated regarding the *travel times* variable, there appears to be an element of artifactual error in the coefficient estimates that can be seen. Also, given the wide MPLE standard errors in the unweighted models, the estimates cannot be viewed as being statistically significant. In the weighted Lorraine models, however, there is a very small increase in the coefficient estimate from the looped models to the non-looped models. This may reflect the fact that once self-transfers are ignored, the remaining inter-hospital distances become larger.

Returning to the hypothesis of MCO length of stay, it was initially thought that two possibilities may occur. Firstly, that hospitals where patients have a longer length of MCO stay would be more likely to transfer patients to SSRs, because patients may have deconditioned through a prolonged hospital stay to the point where it is unsafe for them to be discharged home directly. A longer hospital stay may also indicate a more complex underlying medical condition, or complications having occurred during the MCO stay. However, conversely, it could also be argued that patients may have a shorter MCO stay because they tend to be promptly discharged to SSR hospitals to complete the rest of their recovery. This would be consistent with findings after the 2008 hospital reimbursement reform, where public hospitals becoming more competitive was associated with a shorter length of acute stay (7). In reality, it could be that both of these scenarios involving LOS occur at MCO hospitals, and that the overall statistical effect depends on which of these processes is dominant.

Therefore, it may be that in Bretagne, the former process is dominant, leading to a positive coefficient estimate in the weighted ERGM models, where a longer MCO stay is associated with a greater likelihood of transfer. In Lorraine, the latter process could be driving transfers, where a shorter MCO stay is associated with a greater likelihood of patient transfer in the weighted models. Interestingly, these patterns may also be reflected in the changes to the *no MCO beds* variable. In Bretagne, the *no MCO beds* coefficients increase between models C and D. Recalling that the hospitals with no MCO beds are all SSRs, this may reflect the fact that once we take into account that patients are staying at MCOs for a longer period of time, they are more likely to then be transferred to an SSR. In Lorraine, the corresponding *no MCO beds* coefficients decrease. This could reflect the fact that once patients stay longer at the MCO, they would be less likely to then be sent to SSRs. A possible explanation for these differences in regional patterns could be that in Lorraine, there is more competition between hospitals, given that the proportion of ESPIC hospitals is higher, and that there is a smaller proportion of SSR beds to which patients could be sent.

Finally, in terms of choosing the best model to describe the determinants of inter-hospital patient transfers in the project, the candidates would be either models C or D, as they show the best GOF for the unweighted models. For weighted ERGM, they also perform best in terms of changes to the coefficients for the other variables within the models. Whether C or D is ultimately chosen can be debated. For the unweighted models, it can be seen that model D is not markedly different from model C in terms of changes to variable coefficients, and does not improve GOF (Appendix 8 for example of a GOF plot series). Furthermore, the additional variable *median MCO LOS* is not statistically significant in itself. However, this variable is statistically significant in the weighted models, and it may also be interesting to illustrate the relationship between the variables *median MCO LOS* and *no MCO beds*, as described above, that can be seen in model D.

4.4 Limitations

4.4.1 Limitations in study design

As with all analyses of secondary data, the analyses within this project were limited by the breadth and quality of the information in the dataset, despite the best efforts of the author at crossverification. It cannot, for example, examine other potential determinants of patient transfer, such as the social situation of the patient, or the likelihood of a particular SSR having an unoccupied bed at any given time. As well as this, the analysis only covers one very broad time point in the year of 2012, and does not examine temporal change.

In the process of creating the adjacency matrices for this project, patients being transferred to other regions were also omitted. This may have created boundary effects, especially for hospitals located in the periphery of each region.

4.4.2 Limitations in generalisability

In terms of generalisability, an important limitation is that there have been major changes to many of the administrative structures that underlie the project data, including the 2016 territorial reforms, and changes to the healthcare system (7,11). As well as this, the data used in this project are already several years old. Furthermore, only two specialities were examined. Patients requiring other types of rehabilitation, for example, cardiac or psychiatric, may have different determinants of transfer. All of these factors would mean that these particular results would be less directly useful for any health system planning purposes.

4.5 Recommendations

Having used both unweighted and weighted ERGM techniques with patient transfer data in this project, it was found that both were useful for investigating the determinants of inter-hospital patient transfers. Although both provided similar information, the analysis produced with the weighted technique is perhaps more complete, given that it takes the dimension of weights into account. However, it is also time-consuming, and may not converge for larger networks. For future projects involving similar data or study questions, it may be useful to employ a combination of both techniques, or to begin with the weighted technique and then in the case of non-convergence, to supplement with the unweighted technique. For a more complex network, with a larger number of nodes and edges, it may be necessary to factor in extra time and ensure adequate computing power, as various algorithm control parameters would need to be increased by a large margin, as previously described.

To an extent, the direction of future study in this topic would depend on the ongoing statistical and programming developments. For example, in order to determine an adequate method for assessing GOF for weighted networks, or to be able to analyse categorical edge variables, expansions in the underlying mathematical theory would be required, which would then need to be translated into statistical packages that are compatible with conventional network data formats. In addition, developments in the analysis of bipartite networks may soon allow ERGM methods to be used without the projection technique (39).

Additionally, in terms of considering patient transfer pathways in France, further approaches may involve using patient postcode information to determine whether the residential address of a patient is one of the determinants of transfer and hospital competition (40). Additionally, analyses could be performed after extracting database information and constructing adjacency matrices in a manner that classifies hospitals as only MCOs and SSRs, based on the type of care that was given during a particular hospital stay. The tergm statistical package has now also become available, as an extension of the ergm package. This allows temporal ERGMs to be performed, in order to assess networks at various time points, and this may also be of interest in an evolving healthcare system.

Many conclusions can be drawn on a statistical level through the results of this project. However, in order for this to be useful for health policy, it would be advisable firstly to use more contemporaneous data for analysis, and secondly, to validate some of the study findings through conducting research at the health facilities themselves. This would allow researchers to explore how the data may reflect the day-to-day realities of decision making regarding patient transfers, and to understand the policies, processes and other influential factors at a local level.

5. Conclusion

The field of network theory continues to develop new analytical tools, many of which can be adopted in public health in order to investigate the dynamics within certain systems. In this project, both unweighted ERGM and the newer weighted ERGM technique have been used to investigate the determinants of patient transfers between acute and subacute hospitals in three of the former regions of France. Both techniques gave similar results, but had their own strengths and weaknesses. Notably, there was no way to determine GOF for the weighted model. In general, there also needs to be a better solution for analysing travel times between establishments, especially in the networks without loops. Despite this, the use of the weighted technique, combined with the creation of networks without loops and the inclusion of additional predictor variables, allowed this project to extend the findings from the 2014 FEHAP study. In particular, it confirmed that the regional patient transfer networks are not random, and that the geographic department in which hospitals are situated is an important predictor for patient transfers. It also demonstrated that the legal status of a hospital is statistically significant as a predictor, contrary to the original FEHAP study findings, where it was only significant for Rhône-Alpes (9,15). However, this effect is diminished for the networks without loops, and there is also a relationship between legal status and whether the hospital has MCO or SSR beds. Finally, this project also demonstrates that a range of mechanisms, possibly including the force of competition, may explain the relationship between LOS at MCOs and the likelihood of transfer.

References

1. Ognyanova K. Network Analysis and Visualization with R and igraph. 2016.
2. Levy M. ERGM Tutorial | R-bloggers [Internet]. R-Bloggers. 2016 [cited 2019 Jun 4]. Available from: https://www.r-bloggers.com/ergm-tutorial/
3. Luke DA, Harris JK. Network Analysis in Public Health: History, Methods, and Applications. Annu Rev Public Health. 2007;28:69–93.
4. Mascia D, Di Vincenzo F, Cicchetti A. Dynamic analysis of interhospital collaboration and competition: Empirical evidence from an Italian regional health system. Health Policy (New York). 2012;105:273–81.
5. Mascia D, Angeli F, Di Vincenzo F. Effect of hospital referral networks on patient readmissions. Soc Sci Med. 2015;132:113–21.
6. Weeks WB, Jardin M, Paraponaris A. Characteristics and patterns of elective admissions to for-profit andnot-for-profit hospitals in France in 2009 and 2010. Soc Sci Med. 2015;133:53–8.
7. Choné P. Competition policy for health care provision in France. Health Policy (New York). 2017;121(2):111–8.
8. Ministère des Finances et des Comptes Publics. Décret no 2014-1611 du 24 décembre 2014. Journal Officiel de la République Française 2014 p. 29.
9. Ferrat L. Rapport De Stage De Fin d'Études. La Grande École de la Statistique et du Traitement de l'Information; 2014.
10. Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. maps: Draw Geographical Maps. 2018.
11. Chevreul K, Brigham KB, Durand-Zaleski I, Hernández-Quevedo C. Health Systems in Transition: France Vol. 17 No. 3 2015. Vol. 17, Health system review. 2015.
12. Lee BY, McGlone SM, Song Y, Avery TR, Eubank S, Chang CC, et al. Social network analysis of patient sharing among hospitals in Orange County, California. Am J Public Health. 2011;101(4):707–13.
13. Newman MEJ. The Structure and Function of Complex Networks. SIAM Rev. 2005 Sep 8;45(2):167–256.
14. Brunson JC, Laubenbacher RC. Applications of network analysis to routinely collected health care data: A systematic review. J Am Med Informatics Assoc. 2018;25(2):210–21.
15. Le Meur N, Ferrat L, Gao F, Quidu F, Louazel M. Maillage territorial des établissements de santé : apport des modèles issus de la théorie des graphes. J Gest d'économie médicales. 2017;35(4):197.

16. Dugoff EH, Fernandes-Taylor S, Weissman GE, Huntley JH, Pollack CE. A scoping review of patient-sharing network studies using administrative data. Transl Behav Med. 2018;598–625.

17. Everett MG, Borgatti SP. The dual-projection approach for two-mode networks. Soc Networks. 2013;35:204–10.

18. Landon BE, Keating NL, Barnett ML, Onnela J-P, Paul S, O'Malley AJ, et al. Variation in patient-sharing networks of physicians across the United States. JAMA - J Am Med Assoc. 2012;308(3):265–73.

19. Caimo A, Pallotti F, Lomi A. Bayesian exponential random graph modelling of interhospital patient referral networks. Stat Med. 2017;36:2902–20.

20. Krivitsky PN. Exponential-family random graph models for valued networks. Electron J Stat. 2012;6:1100–28.

21. Ministère des Solidarités et de la Santé. S.A.E Diffusion [Internet]. 2015 [cited 2019 Mar 31]. Available from: https://www.sae-diffusion.sante.gouv.fr/sae-diffusion/accueil.htm

22. Ministère de l'Action et des Comptes publics, Ministère des Solidarités et de la Santé, Ministère du Travail. FINESS : Accueil [Internet]. 2019 [cited 2019 Mar 18]. Available from: http://finess.sante.gouv.fr/fininter/jsp/index.jsp

23. OpenStreetMap Foundation. OpenStreetMap [Internet]. 2019 [cited 2019 Mar 18]. Available from: https://www.openstreetmap.org/about

24. Giraud T, Cura R, Viry M. Package "osrm." 2019.

25. Kolaczyk ED, Csárdi G. Statistical Analysis of Network Data with R. 1st ed. New York: Springer; 2014.

26. Newman MEJ. Mixing patterns in networks. Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top. 2003;67(2):1–13.

27. Farine DR. Measuring phenotypic assortment in animal social networks: Weighted associations are more robust than binary edges. Anim Behav. 2014;89:141–53.

28. Butts CT, Handcock MS, Goodreau SM, Hunter DR, Morris M. ergm : A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks . J Stat Softw. 2015;

29. Krivitsky PN, Handcock MS, Hunter DR. Package "ergm.count." Vol. 6, Electronic Journal of Statistics. 2012. p. 1100–28.

30. Hamra G, MacLehose R, Richardson D. Markov Chain Monte Carlo: An introduction for epidemiologists. Int J Epidemiol. 2013;42:627–34.

31. Liu J, Nordman DJ, Meeker WQ. The Number of MCMC Draws Needed to Compute Bayesian Credible Bounds. Am Stat. 2016 Jul 2;70(3):275–84.

32. Geweke J. Getting it right: Joint distribution tests of posterior simulators. J Am Stat Assoc. 2004 Sep;99(467):799–804.

33. Krivitsky PN, Butts CT. Modeling valued networks with statnet. 2015. p. 1–58.

34. Morris M, Handcock MS, Hunter DR. Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. J Stat Softw. 2008;

35. Duxbury SW. Diagnosing Multicollinearity in Exponential Random Graph Models. Sociol Methods Res. 2018;1–40.

36. Snijders TAB. Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. J Soc Struct. 2002;2.

37. Shore J, Lubin B. Spectral goodness of fit for network models. Soc Networks. 2015 Oct 1;43:16–27.

38. Li Y, Carriere KC. Assessing Goodness of Fit of Exponential Random Graph Models. Int J Stat Probab. 2013 Oct 11;2(4):64–74.

39. Wang P, Sharpe K, Robins GL, Pattison PE. Exponential random graph (p *) models for affiliation networks. Soc Networks. 2009;31:12–25.

40. Garnick DW, Luft HS, Robinson JC, Tetreault J. Appropriate measures of hospital market areas. Health Serv Res. 1987;22(1):69–89.

41. Agence Technique de L'information sur L'hospitalisation. Manuel des GHM - Version 11d [Internet]. 2013 [cited 2019 May 19]. Available from: https://www.atih.sante.fr/manuel-des-ghm-version-11d

42. Handcock M, Hunter D, Butts C, Goodreau S, Krivitsky P, Morris M. Statnet suite of packages. 2018.

43. Wickham H, Henry L. tidyr: Easily Tidy Data with "spread()" and "gather()" Functions. 2019.

44. Farine D. assortnet: Calculate the Assortativity Coefficient of Weighted and Binary Networks. 2016.

45. Fox J, Weisberg S. An {R} Companion to Applied Regression. 2011.

46. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. 2019.

47. Csardi G, Nepusz T. The igraph software package for complex network research. 2006.

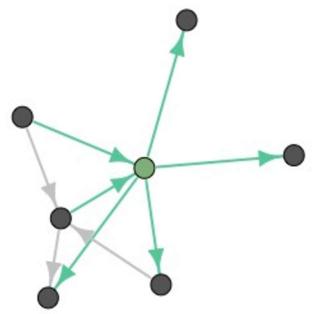
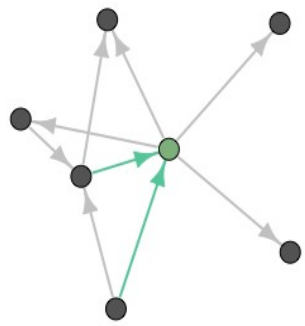
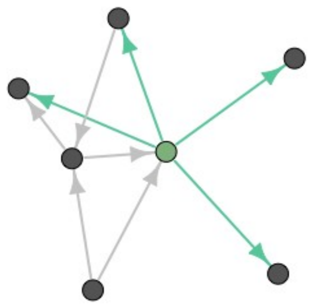
48. Bojanowski M. intergraph: Coercion Routines for Network Data Objects. 2015.

Appendices

Appendix I: Descriptive network statistics

In the field of network theory, occasionally more than one definition is used for a given measure. Where this is the case, only the definition used for the analysis of the directed graphs within this project is included.

Table i: Descriptive network statistics

Statistic	Definition	Diagram
Degree	Applied to a node: The total number of edges connected to a particular node. The arrows in green in the diagram are the edges that add up to the degree of the node in green. Therefore, the degree of this node is 6. Applied to the network: The mean degree is usually calculated. This is the mean number of edges for each node in the network.	
- In-degree	Applied to a node: The total number of edges travelling to a particular node. The arrows in green in the diagram are the edges that add up to the in-degree of the node in green. Therefore, the in-degree of this node is 2. Applied to the network: The mean in-degree is the mean number of edges travelling to each node in the network.	
- Out-degree	Applied to a node: The total number of edges leaving from a particular node. The arrows in green in the diagram are the edges that add up to the out-degree of the node in green. Therefore, the out-degree of this node is 4. Applied to the network: The mean in-degree is the mean number of edges leaving from each node in the network.	
Density	Statistic applied to the entire network: The number of edges there are in the network, as a proportion of the total number of possible edges. This measure is unweighted, meaning that it does not take into account the weight of each of the observed edges.	
Assortativity	The similarity, in terms of a particular attribute, of the two nodes at the ends of each edge, as summed for the entire network. (See 2.5 for a more detailed description.)	

Appendix 2: Variables extracted from PMSI and SAE databases

From the PMSI database, the following information was obtained for individual patients:

- Unique patient ID number, as assigned by the healthcare system: This number is used to match each patient as they are transferred from one hospital to another.
- Age
- Sex
- FINESS number of the facility where the stay occurred
- Unique hospital stay ID number, as assigned by the healthcare system: This number assisted with separating out different stays that the patient may have had at the same hospital during the year.
- Length of stay (LOS) for each hospital stay: Therefore, for a patient who was initially at an MCO and then transferred to an SSR, there would be a LOS value for the MCO stay, and a LOS value for the SSR stay.
- Speciality of care: This was coded according to the *Catégorie majeure de diagnostic* (major diagnostic category) version 11d for MCO stays, and *Catégorie majeure clinique* (major clinical category) version 7 for SSR stays (41).
- Origin: This denotes whether the patient entered the hospital from home, or by transfer from another health facility.
- Mode of entry: This is another parameter that allows specification of whether a patient entered a particular episode of care as an admission from the community, as a transfer from another health facility, or due to a change in clinical teams within the same hospital.
- Destination: This denotes whether a patient was discharged home from hospital, or transferred to another health facility.
- Mode of exit: This is another parameter that allows specification of whether an episode of care ended because a patient was discharged, transferred, had died during the stay, or changed clinical teams within the same hospital.

The hospital-level parameters extracted from the SAE database were as follows:

- FINESS number for each hospital: This is a unique hospital identifier code, and was used to match stays that took place within the same health establishment. Some hospitals also have a judicial FINESS number, used to identify a larger business entity under which the hospital belongs, and this was at times a source of inaccuracies in the data (9).
- Commune code: This is the smallest geographical administrative area of each hospital, and allows travel times between hospitals to be calculated.
- Department code: This is a larger geographical area compared to a commune, and allows sub-grouping of hospitals by location within a region.
- Region: This is the largest geographical unit considered in this project, and is demarcated according to the regions of France in 2012, prior to the 2016 national territorial reforms.

- Class: This denotes the type of care offered by the hospital, according to the categories described above of MCO, SSR or mixed MCO/SSR.
- Legal status: The sector to which a hospital belongs, in terms of public, private not-for-profit (ESPIC) or private for-profit.
- Number of MCO beds: This number was obtained by adding the number of acute medical and surgical beds reported by the hospital, while ignoring the number of obstetric beds, as a high number of obstetric beds may artificially inflate the size of a hospital, while being irrelevant for the specialties being examined in this study.
- Number of SSR beds: This number was obtained from the reported number of adult SSR beds only.

Appendix 3: R packages used

R versions used in project:

- 3.5.1 "Feather Spray"
- 3.5.2 "Eggshell Igloo"
- 3.6.0 "Planting of a tree"

RStudio versions used in project:

- 1.1.463
- 1.2.1335

R packages used in project:

Table ii: R packages used in project

Package	Brief description	Version
assortnet	Calculates weighted assortativities and their standard errors (44)	0.12
car	Allows the calculation of VIF for linear regression models (45)	3.0-2
dplyr	A set of tools for data manipulation (46)	0.8.0.1
ergm	Performs unweighted ERGMs and a range of related functions (28)	3.9.4
ergm.count	Performs weighted ERGMs and a range of related functions (29)	3.3.0
igraph	Allows manipulation and analysis of networks (47)	1.2.4.1
intergraph	Converts between networks created in igraph and statnet (48)	2.0-2
maps	Allows visual representation of geographical locations (10)	3.3.0
osrm	Retrieves travel times between locations (24)	3.2.0
statnet	Allows manipulation and analysis of networks (42)	2018.10
tidyr	A set of tools for data manipulation (43)	0.8.3

Appendix 4: Original FEHAP Study model specifications

The following models were built for each of the 3 regions in the FEHAP study (9).

Table iii: Model specifications, FEHAP Study 2014

	Model 1	Model 2	Model 3	Model 4
Edges	✓	✓	✓	✓
Legal status	✓	✓	✓	✓
MCO-MCO		* Constrained to -infinity		* Constrained to -infinity
SSR-SSR		* Constrained to -infinity		* Constrained to -infinity
BedSSR			✓	✓
BedMCO			✓	✓
Time			✓	✓
Department			✓	✓
Isolates			✓	✓

Key: Shaded in grey = not included in model; Tick = included in model

Appendix 5: Mixing matrices for assortativity by legal status

Table iv: Mixing matrices for assortativity by legal status, all regions

	Bretagne		Lorraine		Rhône-Alpes	
	With loops	No loops	With loops	No loops	With loops	No loops
Assortativity by legal status						
Estimate	**0.25**	0.05	**0.31**	0.06	**0.28**	0.08
Standard error	0.06	0.05	0.07	0.05	0.05	0.04
Public/public	0.4635	0.1902	0.4324	0.3015	0.3985	0.1913
Public/private	0.0789	0.1248	0.0405	0.0627	0.1140	0.1619
Public/ESPIC	0.1647	0.2604	0.1950	0.3019	0.1223	0.1738
Private/public	0.0661	0.1044	0.0741	0.1147	0.0733	0.1042
Private/private	0.0719	0.1137	0.0311	0.0234	0.1096	0.1461
Private/ESPIC	0.0947	0.1497	0.0840	0.1300	0.07476	0.1062
ESPIC/public	0.0040	0.0064	0.0072	0.0112	0.0173	0.0245
ESPIC/private	0.0159	0.0251	0.0011	0.0017	0.0294	0.0418
ESPIC/ESPIC	0.0402	0.0253	0.1346	0.0530	0.0609	0.05027

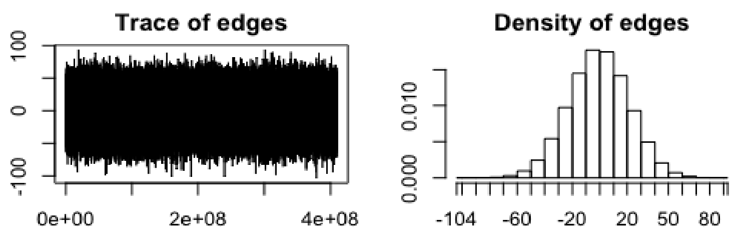
Appendix 6: Example of R code for weighted ERGM, Lorraine model D

```
## WEIGHTED ERGM Model D: Lorraine (LOR) WITH Loops ##


load("~/grapheLOR_May15.RData")

load("~/netLOR_May15.RData")

load("~/rdiagmtxLOR_May20.RData")


LORModDW <- ergm(netLOR ~ sum + nodematch("legalstatus") + nodematch("department")

            + nodecov("MCObeds") + nodecov("SSRbeds")

            + nodecov("noMCObeds") + nodecov("noSSRbeds")

            + nodecov("medianMCOLOS")

            + edgecov(rdiagmtxLOR[V(grapheLOR)$name,V(grapheLOR)$name])

            + nodemix("class", base=c(-1, -2, -3, -6, -9)),

            response="weight", reference=~Poisson,

            control=control.ergm(drop=TRUE,MCMC.samplesize=2e+5,MCMLE.maxit = 60,
                MCMC.interval=10000,  MCMC.burnin = 2e+5, MCMLE.trustregion=1000,
                MCMLE.steplength=1, force.main=TRUE))

proc.time()

save(LORModDW, file="LORModDW.RData")

summary(LORModDW)


# Checking MCMC diagnostics

par(mar = rep(2,4))

mcmc.diagnostics(LORModDW)
```

Appendix 7: Example of MCMC diagnostic plots generated

Due to the number of data points that are plotted, and the file sizes in which this results, I have only included the MCMC plots of 1 variable from an earlier ERGM run where “only” 4 x 10^8 Markov chain proposals were requested.

Fig i: Example of MCMC diagnostic plots generated by the ergm package



Appendix 8: Progression of GOF plots, unweighted Bretagne model with loops

Fig ii Model A Left plot: x axis = In-degree, y axis = proportion of nodes; Right plot: x axis = Out-degree, y axis = proportion of nodes

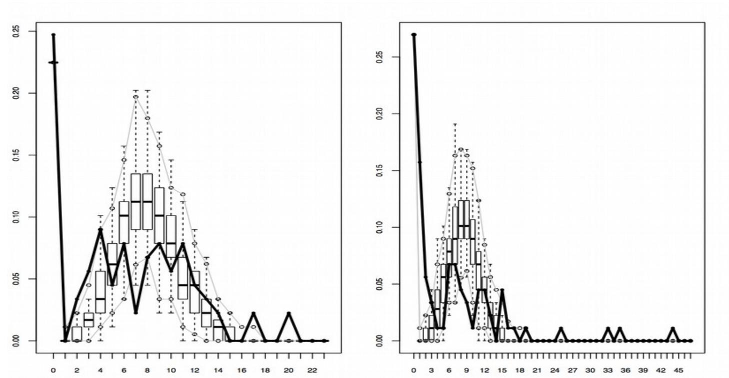


Fig iii Model B Left plot: x axis = In-degree, y axis = proportion of nodes; Right plot: x axis = Out-degree, y axis = proportion of nodes

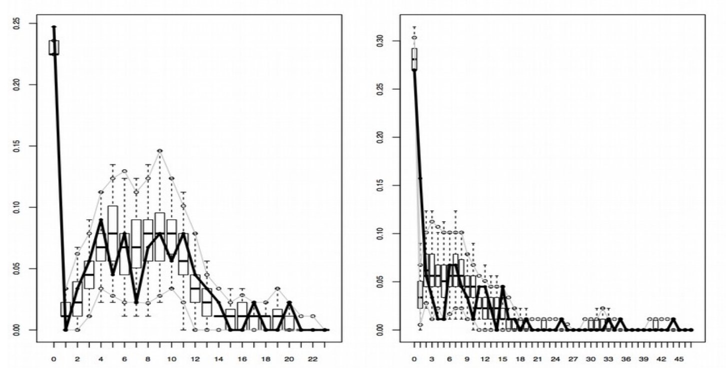


Fig iv Model C Left plot: x axis = In-degree, y axis = proportion of nodes; Right plot: x axis = Out-degree, y axis = proportion of nodes

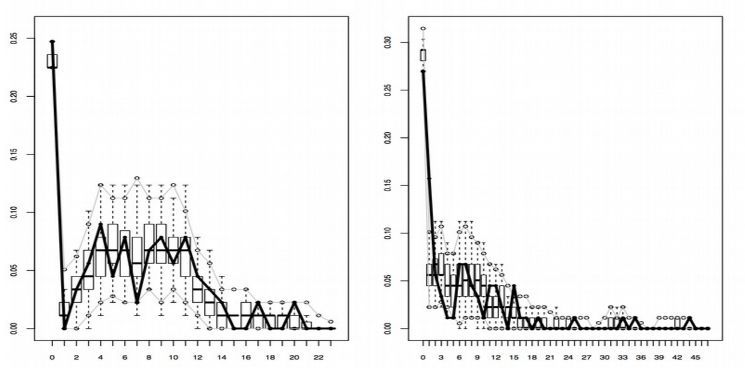
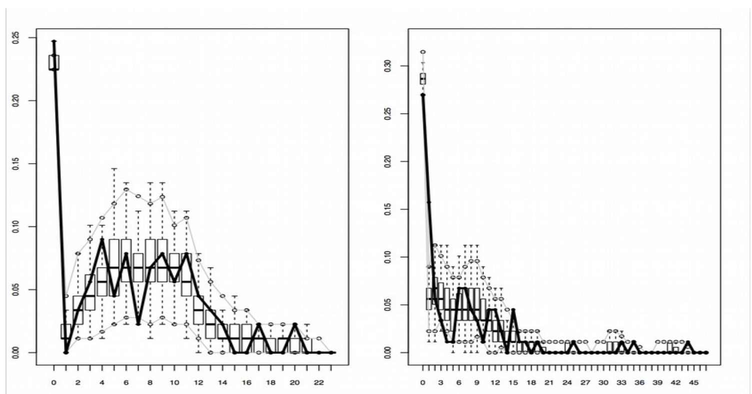


Fig v Model D Left plot: x axis = In-degree, y axis = proportion of nodes; Right plot: x axis = Out-degree, y axis = proportion of nodes



Appendix 9: Computation times for weighted ERGM models

Computers used:

- “Cluster”: GenOuest Bioinformatics Core Facility cluster, up to 50GB memory (variable depending on other users)
- “Linux”: DELL with Intel Processor with 12 cores, 32GB memory
- “Mac”: 2017 MacBook Pro with dual core, 3.1GHz processor, 8GB memory

Table v: Computation times for weighted Bretagne models

Bretagne – Model size: 89 nodes, 556 edges (with loops), 513 edges (without loops)								
	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Computer	Linux	Mac	Linux	Cluster	Mac	Cluster	Mac	Cluster
Iterations	37	27	33	34	35	36	36	35
Run-time (h)	33	20	47	33	22.5	32.5	38	34.5

Table vi: Computation times for weighted Lorraine models

Lorraine – Model size: 80 nodes, 513 edges (with loops), 477 edges (without loops)								
	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Computer	Cluster	Mac	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster
Iterations	26	18	22	21	23	23	24	23
Run-time (h)	24	10	5.5	24	6	31	8	33

Table vii: Computation times for weighted Rhône-Alpes models

Rhône-Alpes – Model size: 199 nodes, 1594 edges (with loops), 1524 edges (without loops)								
	A		B		C		D	
	Loops	No loops	Loops	No loops	Loops	No loops	Loops	No loops
Computer	Mac	Mac	Cluster	Cluster	Mac	Cluster	Mac	Cluster
Iterations	34	22	65*	68	65*	75*	65*	75*
Run-time (h)	18.5	12	84	8	72	9.5	88	63.5

*Key: * = Non-convergence by the stated number of iterations*

Appendix 10: Ethics documentation – Secondary Data Declaration and Approval to use PMSI Data

Department of Economics/ScHARR

Research Ethics Review
for Undergraduate and Postgraduate-Taught Students

Form 1C: Student Declaration

To be included in Appendices of dissertation

→ **Research Project Title:** Deciphering Hospital Networks Using Graph Theory Methods

→ **Name of dataset to be used:** SAE (Statistiques annuelles des établissements de santé

Owner of dataset: Ministère des Solidarités et de la Santé, France

Total number of datasets to be used: ....1.... If more than one, then fill in a separate declaration for each dataset.

State the case that applies to your research project: Case ...1.....

Case 1: Your proposed project will only involve anonymised/aggregated data that any member of the public is legitimately free to access and use without having to obtain permission from anyone else. E.g., macroeconomic statistics provided by legitimate sources such as government departments and international organisations; anonymised secondary data on individuals or firms provided by legitimate sources such as government departments and which do not require any form of registration or statement of purpose to allow access.

Case 2: Your proposed project will involve anonymised secondary data for which you need to obtain permission from the owner (e.g., you need to satisfy some condition before being permitted to download the data, such as a declaration of intended educational purpose. Downloading the BHPS from the Data Archive falls in this category.)

Case 3: None of the above cases. Note that the department does not allow undergraduate or postgraduate taught students to use primary data, or secondary data that may include personal data, unless specific training is undertaken by the student.

→ If your proposed project falls within Case 1, then simply print your name, date and sign below.
If your proposed project falls within Case 2, then you need to append to this form evidence that you have legitimately obtained access to these data. E.g.,

Department of Economics, / ScHARR July 2011.

confirmation email, and statement of purpose if one was required. Then print your name, date and sign below.
If your proposed project falls within Case 3, then contact your supervisor or supervisory team as soon as possible. You may not be able to use the proposed data.

→ **Name of student:** Yuanfei Anny HUANG

→ **Signature of student:** 黄 Date: 14/03/2019

→ **Name of supervisor:** LE MEUR Nolwenn

→ **Signature of Supervisor:** Date: 20/03/2019

Department of Economics/ScHARR

Research Ethics Review
for Undergraduate and Postgraduate-Taught Students

Form 1C: Student Declaration

To be included in Appendices of dissertation

Research Project Title: Deciphering Hospital Networks Using Graph Theory Methods

Name of dataset to be used: PMSI (Programme de Médicalisation des Systèmes d'Information)

Owner of dataset: (CNIL) Commission Nationale de l'Informatique et des Libertés

Total number of datasets to be used: ....2.... **If more than one, then fill in a separate declaration for each dataset.**

State the case that applies to your research project: Case ...2.....

Case 1: Your proposed project will only involve anonymised/aggregated data that any member of the public is legitimately free to access and use without having to obtain permission from anyone else. E.g., macroeconomic statistics provided by legitimate sources such as government departments and international organisations; anonymised secondary data on individuals or firms provided by legitimate sources such as government departments and which do not require any form of registration or statement of purpose to allow access.

Case 2: Your proposed project will involve anonymised secondary data for which you need to obtain permission from the owner (e.g., you need to satisfy some condition before being permitted to download the data, such as a declaration of intended educational purpose. Downloading the BHPS from the Data Archive falls in this category.)

Case 3: None of the above cases. Note that the department does not allow undergraduate or postgraduate taught students to use primary data, or secondary data that may include personal data, unless specific training is undertaken by the student.

If your proposed project falls within Case 1, then simply print your name, date and sign below.

If your proposed project falls within Case 2, then you need to append to this form evidence that you have legitimately obtained access to these data. E.g., confirmation email, and statement of purpose if one was required. Then print your name, date and sign below.
If your proposed project falls within Case 3, then contact your supervisor or supervisory team as soon as possible. You may not be able to use the proposed data.

→ **Name of student:** Yuanfei Anny HUANG

→ **Signature of student:** 黄艾 Date: 14/03/19

→ **Name of supervisor:** LE MEUR NOLWENN

→ **Signature of Supervisor:** Date: 20/03/2019

Department of Economics, / ScHARR July 2011.

\

CNIL

Le Vice-Président délégué

COURRIER ARRIVÉ
- 5 AVR. 2012
DIRECTION

Monsieur Antoine FLAHAULT
DIRECTEUR
ECOLE DES HAUTES ETUDES EN SANTE PUBLIQUE (EHESP)
AVENUE DU PR LEON BERNARD
35043 - RENNES

A l'attention de Mme Sahar BAYAT

Paris, le 30 MARS 2012

N/Réf. : EGY/VCS/AE121029

Objet : NOTIFICATION D'AUTORISATION

Décision DE-2012-032 autorisant l'ECOLE DES HAUTES ETUDES EN SANTE PUBLIQUE à mettre en œuvre un traitement de données de santé à caractère personnel ayant pour finalité la réalisation d'études visant à évaluer les pratiques et les trajectoires de soins, à partir des données PMSI-champs MCO, SSR, HAD, PSY avec chaînage ANO, à partir de l'année 2006. (Demande d'autorisation n° 1564135)

Monsieur le Directeur,

Vous avez saisi notre Commission d'une demande d'autorisation relative à un traitement de données à caractère personnel ayant pour finalité :

EVALUER LES PRATIQUES ET LES TRAJECTOIRES DE SOINS, A PARTIR DES DONNEES PMSI-CHAMPS MCO, SSR, HAD, PSY AVEC CHAINAGE ANO, A PARTIR DE L'ANNEE 2006.

Ce traitement relève de la procédure des articles 62 et suivants de la loi du 6 janvier 1978 modifiée.

Vous indiquez que des mesures de sécurité physique et logique seront mises en place pour garantir la confidentialité des données et que le traitement informatique des données sera réalisé sous votre responsabilité et celle de vos collaborateurs.

J'attire votre attention sur les obligations qui incombent désormais à ces personnes qui doivent :

- n'utiliser les fichiers qu'à des fins d'analyse comparative de l'activité de soins,
- respecter et faire respecter le secret des informations cédées par toutes les personnes susceptibles de travailler sur ces données, ces personnes étant astreintes par écrit au secret professionnel,
- prendre toutes précautions utiles afin de préserver la sécurité des informations ainsi transmises et notamment empêcher qu'elles ne soient déformées, endommagées ou communiquées à des tiers non autorisés,
- ne pas rétrocéder ou divulguer à des tiers les informations fournies sous quelque forme que ce soit,
- ne pas procéder à des rapprochements, interconnexions, mises en relation, appariements avec tout fichier de données directement ou indirectement nominatives ou toute information susceptible de révéler l'identité d'une personne et/ou son état de santé,
- ne pas utiliser de façon détournée les informations transmises, notamment à des fins de recherche ou d'identification des personnes.

Commission Nationale de l'Informatique et des Libertés
8 rue Vivienne CS 30223 75083 PARIS Cedex 02 - Tél: 01 53 73 22 22 - Fax: 01 53 73 22 00 - www.cnil.fr