

Social network analysis applied to healthcare data: inventory, discussion and perspectives

Coralie Gandré

► **To cite this version:**

Coralie Gandré. Social network analysis applied to healthcare data: inventory, discussion and perspectives. 2019. hal-02054253

HAL Id: hal-02054253

<https://hal.archives-ouvertes.fr/hal-02054253>

Submitted on 1 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Social network analysis applied to healthcare data: inventory, discussion and perspectives

Summary of the ARAQNEE workshop

Coralie Gandré¹

Institut de recherche et documentation en économie de la santé

Submitted on January 30th 2019, published on March 1st 2019

Introduction

Social network analysis (SNA) is a set of methods and concepts based on graph theory that analyse relational phenomena. SNA has been developed in fields ranging from social science to exact science, and can have numerous applications in health research. It can for instance be used to study collaborative networks of health professionals, care pathways, the spread of infectious diseases and risky behaviours, and the dissemination of health prevention and promotion programs. In particular, the recent availability of health claim data for research purposes, such as the French National Health Data System (*Systeme national des données de santé*, SNDS) provides multiple opportunities for further SNA-based studies.

At present, the application of SNA for the study of health-related issues is limited in France in comparison with other Western countries, despite its expected benefits. There is a growing need for developing synergies between teams currently applying SNA in the field of health research, to encourage the use of this method and to highlight the potential benefits of SNA to support decision-making in healthcare.

On the 28th November 2018, a one-day workshop was held in Paris to present an inventory of research applying SNA to healthcare data in France and other countries and to provide methodological insights from international keynote speakers (*Journée d'étude Analyse de Réseaux*

¹ gandre@irdes.fr

AppliQuée aux donNÉes de Santé, Araqnée). This workshop was organized by the French Institute for Research and Information in Health Economics (*Institut de recherche et documentation en économie de la santé, Irdes*), in collaboration with the French School of Public Health (*École des hautes études en santé publique*”, EHESP) and the International College of Territorial Science (*Collège international des sciences territoriales*”, Cist). It benefited from financial support from the French Institute for Public Health Research (*Institut de recherche en santé publique, Iresp*) and from the Research Network for Network Analysis in Human and Social Sciences (*Groupement de recherche analyse de réseaux en SHS, GDR 3771*). The sessions featured presentations by researchers, including both senior and junior professionals, as well as open discussions with the audience and a general discussion led by a dedicated discussant. The audience included researchers from diverse fields (including public health, geography, economy, sociology and mathematical science), decision-makers, health professionals and health service users. This article provides a brief overview of the contents of the workshop and presents elements of discussion and future perspectives.

Contents of the workshop

Application of SNA to the study of healthcare provider networks using health claim data

For several decades, research has showed wide variations in patterns of healthcare delivery, which cannot be explained by the patients’ clinical and socio-demographic characteristics nor by their individual preferences in terms of care. Local medical practice style plays a significant role in these variations and is likely to be shaped by physicians’ social interactions. The study of networks of health professionals, whether formal (such as accountable care organizations – ACOs in the US) or informal, and their impact on care delivery can therefore provide key insights to better understand variations in care provision. Previous research has highlighted the importance of understanding and characterizing the structure of professional networks to ensure an optimal quality of care (Cunningham *et al.*, 2012). Limitations in data collection have long been an obstacle: SNA has often

been based on survey data which restricted the size and impact of studies as such surveys often have a low response rate and are expensive and resource-intensive to administer (Barnett *et al.*, 2011). With the increased availability of health claim data for research purposes and the emergence of new approaches from complex network studies (physics, computer sciences), SNA shows promise of being developed on a larger scale and at a lower cost using health claim data. Empirical research has demonstrated that the sharing of patients between health professionals predicts significant relationships between these professionals (Barnett *et al.*, 2011). This approach was followed by four presentations during the workshop.

The first, by Bruce Landon (Harvard medical school, United States), focused on patient sharing networks in the US as a lever to improve healthcare delivery. His team used claim data from Medicare, a federal health insurance scheme mostly targeting people aged 65 or over, concerning all patients living in 51 randomly sampled hospital referral regions over a five-year period. Landon first built a bipartite network where the nodes were patients and physicians, and the edges were the contacts extracted from the claim data. He then produced a unipartite projection of the bipartite network by directly connecting each pair of physicians who shared patients. Edges between them were weighted by the number of shared patients (Landon *et al.*, 2012). Naturally occurring communities of tightly linked physicians were also identified within the networks. The different communities were described using complementary indicators that characterize coordination, including the percentage of primary care physicians in the community: degree (number of other physicians to whom a physician was connected); and physician dispersion (extent to which a physician's patients were treated by physicians within the community). The relationship between these networks' properties and outcomes was assessed using regression models. Network measures were significantly associated with total annual spend per patient and utilization of services (for example the mean number of specialty office visits or emergency visits). The various measures of quality of care (such as readmissions within 30 days of discharge, mammography screening or regular low-density lipoprotein testing for patients with diabetes or cardiovascular diseases) were

inconsistently related to network measures (Landon *et al.*, 2018). The research team also used network analyses to try to explain the disparities observed between black and white patients (such as the fact that the physicians whom the black patients visited were less likely to be Board certified or that they faced greater difficulties in obtaining access to high-quality specialized care for their patients) (Bach *et al.*, 2004). Physician networks were constructed based on visits by black patients, white patients or both, and sharing patterns were then studied to assess if they differed by race. Findings indeed showed significant variations in network structure by race.

The second presentation, by Coralie Gandré (Institute for research and information in health economics, France), focused on the study of care coordination in psychiatry in the Provence-Alpes-Côte d'Azur (PACA) region using network analysis applied to SNDS claim data. The methodology was similar to that of the first presentation. First, a bipartite network was built where the nodes were patients hospitalized at least once for a severe mental disorder over a two-year period as well as health professionals or hospitals involved in the follow-up of mental disorders one year after the index hospitalization. Secondly, a unipartite network was built where the nodes were health professionals or hospitals and the edges were formed by shared patients and weighted by their number. The network analysis was performed on the ego-networks of index admission hospitals, due to the fact that the study focused on patients with severe mental disorders hospitalized at least once. The network measures included were those most likely to reflect different models of care coordination (density, transitivity, number of nodes and edges, proportion of the different types of health professionals in the network, mean number of annual care contacts per patient of the network). Principal component analysis and agglomerative hierarchical clustering were carried out to provide a typology of the care coordination models of the different ego-networks. 48 ego-networks were built across the PACA region. These can be classified into seven clusters, which present different structural characteristics: some – but not all – are specific to a type of hospital while the size of the hospital and its geographical environment also play a role. The next step of this project will be to link these findings to the

quality of care measured across the different networks and to identify network structures linked to optimal patient outcomes.

The third presentation, by Dominik von Stillfried (Central institute for social health insurance physician care, Germany), focused on the identification of networks of physicians caring for patients with depression in Germany. This work relied on the assumption that while patients may access any number of physicians in the country without financial disincentive or compulsory registration with a general practitioner (GP), most patients are cared for by a virtual network of physician practices, and their outcomes can be attributed to this network rather than to individual practices. Contacts between patients and physicians were extracted from the dataset of German national physician claims for ambulatory care. Each patient was then allocated to the GP who provided the majority of his/her care, and this patient's contacts with other physicians were aggregated around this GP so that all the physicians seen by the patient were allocated to the GP's network. Community detection was also carried out within the networks. While similarities were found in the networks' structures (in particular, most networks did not include relevant specialists such as psychiatrists or psychotherapists, which questions the quality of referral patterns), some variations were also found across networks (in particular in terms of the average number of physicians per network). A comparison between networks based on patients suffering from other disorders also showed that diseases seem to drive network structure, even within the same region. While physicians can be part of several networks, each network appears to treat a specific patient population. The method makes it possible to compare the networks in terms of their outcomes for their patient population, by matching claim data with clinical data or patient-reported outcome measures. This has implications for explaining geographic variations and designing targeted interventions in networks with outliers.

The fourth presentation, by Nolwenn Le Meur (School of public health, France), focused on the applications of graph theories for the analysis of healthcare data. It described in particular a research project that studied the usefulness of resorting to graph theory methods to model transfers of patients between short-stay hospitals (excluding

psychiatry) and follow-up care and rehabilitation facilities. Hospital discharge data, which is now part of the SNDS database, was used for two major groups of clinical diagnosis – musculoskeletal system and neurologic disorders – in three French regions (Bretagne, Lorraine and Rhône-Alpes). To test the significance of assortativity (the preference for a network's nodes to be connected with other nodes that are similar), Erdős-Renyi and constrained degree sequence models were carried out. Findings showed the propensity of healthcare facilities with the same legal status to transfer patients to each other. Block models were then conducted to build clusters of healthcare facilities with common features; they showed the significance of territorial dynamics in the transfers between establishments. Finally, exponential random graph models (ERGM) were performed to identify determinants of network topology. Although these confirmed the influence of geographical proximity and legal status in the relationships between hospitals, regional specificities were also identified, which were probably linked to population characteristics and to the historical structuration of health services in each region (Le Meur *et al.*, 2017).

Others applications of SNA

The application of SNA to healthcare data has not been limited to the use of medical claim data by public health researchers. French teams presented illustrative examples showing that health geographers and sociologists also use these methods to study health services.

The fifth presentation, by Bertrand Lefebvre (School of public health, France), analysed the spatial diffusion of formal private hospital networks (hospital chains) in India. The aim of the research project presented was to study the diffusion of these hospital chains across India and their potential contribution to the reduction of territorial inequalities in the supply of hospital care. Lefebvre used percolation theory to account for the heterogeneity of the diffusion medium (in particular variations in the hospital chains' size and development strategies). To represent the diffusion environment and the links between the sites of the different hospital chains, percolation networks were created by using a database presenting the hospitals according to the network to which they belong as well as a geographic information

system on India's urban system. Several percolation networks were built according to urban hierarchy levels to account for the different strategies (metropolitan, regional or local) developed by the hospital networks and their aggregated effect on the spatial diffusion of the sector (Lefebvre, 2013). Coupling percolation networks and urban hierarchy provides a powerful visualization of the spatial diffusion of hospital chains and helps measure the extent of the sector's regionalization and of the local diffusion processes, as well as the magnitude of certain clusters. This methodology can be applied to other health networks to reveal unexpected spatial location strategies.

The sixth presentation, by Elise Autrive (Rouen university, France), focused on the territorialization of mobile health and social care for vulnerable populations (including populations affected by precarity and instability, addictions, difficulty of communication or comprehension, exposure to direct or indirect violence and lack of access/rights to social and health protection) in French Guiana. The objectives were to identify these services' distribution patterns and spatial location strategies. Semi-structured interviews were conducted with the mobile services' medical and social workers, with a focus on operators' activities, the types of relationships between operators and their geographical mobility. A network was then created: the nodes were operators, with attributes relating to their domain of activity and their geocoded location, and the edges were the interactions (operational or strategic) between nodes. The network analysis demonstrated that the mobile health and social care supply for vulnerable populations was developed by focusing on thematic complementarity with other operators and the needs of vulnerable populations, while being influenced by urban policies. This analysis therefore demonstrated added value in comparison to more common mapping approaches.

The last presentation, by Benoit Cret and Jaime Montes-Lihn (Université Jean-Moulin Lyon 3, France), focused on a multi-level analysis of coordination among primary care teams combining medical and paramedical activities in France (*maisons de santé pluri-professionnelles*, MSP). This project was based on the theoretical frameworks of sociology of work, sociology of organizations and network sociology (Bergeron and Castel, 2010; Lazega, 2001; Pinell,

2005). It aimed to identify and explain cooperation issues in the context of MSPs, by considering the organizational level (formal communication between MSPs and other local health organizations) and individual level (informal communication between individuals working in these different organizations), as well as interactions between these two levels. The data collection method relied on a survey involving MSPs and other local health organizations, aimed at obtaining information both on individuals (socio-demographic information, professional experience, role in the organisation, formal commitment in other organizations, etc.) and on organizations (number of professionals involved, legal status, organizational practices, etc.). The authors analysed the data using multilevel ERGM to understand general and structural trends in the network, associated with social mechanisms underlying collective action in an inter-organizational setting. Preliminary findings showed that the organizational level had its own logic, which was in particular influenced by the size of the organization. The individual level also appeared to have its own logic: the network's structure (in particular in terms of reciprocity, centrality and structural equivalence) impacted whom the individual chose to interact with. The existence of social nests – defined as links between structurally equivalent individuals – was also confirmed within the network. These were influenced both by profession (for instance, individuals from the same profession tend to avoid each other) and by socio-demographic characteristics (for instance, individuals who present similarities along these characteristics tend to interact with each other more), but not by the individual's status within the organization (in particular representatives of healthcare organisations do not tend to ask each other for advice). Finally, little interdependence was shown between levels.

Discussion and perspectives

The workshop highlighted a quickly developing area of research. The application of network analysis to health claim data, inspired by important international work, is being increasingly implemented in France to study the impact of healthcare provider networks on care delivery in an exploratory fashion. Other research emanating from

French teams also uses SNA and survey data within various conceptual frameworks and research areas such as health geography and organizational sociology. However, for feasibility reasons, the selection of research projects to be presented during the workshop was designed to provide illustrative examples of current research projects applying SNA to healthcare data in France: the workshop is therefore not an exhaustive inventory of all currently ongoing projects. In particular, there was a strong focus on health service research while SNA can be used in other areas of public health, such as the study of the spread of infectious diseases (Nekkab *et al.*, 2017) or the development of health behaviour interventions (Shelton *et al.*, 2019).

Current methodological issues

The research projects and discussions raised common methodological issues, which are not entirely resolved. First, there remain questions related to the identification of meaningful networks when using health claim data, in particular whether or not these networks should be built by focusing on specific disorders. In addition, there is no consensual method to determine the threshold of shared patients that can be considered as a marker of actual collaboration between healthcare providers. Most research used an empirical approach by testing several thresholds. Current methodological improvements include the restriction of linkages to healthcare providers who share patients with one another during an episode of care: this makes it possible to exclude potentially spurious links between healthcare providers who treat the same patients for completely unrelated conditions (Landon *et al.*, 2018).

Secondly, the analysis of healthcare provider networks raises some methodological issues. Most studies only use a few network measures among all those available (such as for example degree or density). This has also been noted in other fields of health research (Shelton *et al.*, 2019). It is indeed difficult to find measures that can adequately characterize coordination within healthcare provider networks. While those that are currently used can help characterize different models of care coordination, it remains difficult to draw conclusions on the quality of coordination using such measures. There should be a strong focus on linking network measures with patient outcomes, in order to identify

the best models of coordination. In addition, the transformation of bipartite networks into unipartite networks might require a stronger preliminary focus on the description of bipartite networks as this transformation is associated with information loss.

Thirdly, research applying SNA to health claim data needs to take into account the limitations inherent to the use of such data. Researchers are restricted to the information available in the databases. Very often, these only include healthcare encounters and do not provide any information on health and social care or social care. Similarly, these databases usually do not make it possible to identify the individual health professionals working in hospitals, which constrains the conduction of multi-level analyses. Additionally, while the use of shared patients to identify healthcare provider networks has been validated (Landon *et al.*, 2013), it might be useful to complement it with other indicators, such as the existence of written or email communications between health professionals which are generally not reported in claim data. It appears necessary to pair macro quantitative approaches with more sociological or local approaches, such as the studies presented in the last three contributions of the workshop, to ensure a thorough understanding of the mechanisms at play (for instance whether a network is built upon physician or patient behaviours). However, mixed methods are still only scarcely used in recent international work (Gandré, 2018). Finally, accessing health claim data remains a challenge, which might limit the number of researchers using this data.

These methodological issues in the application of SNA to healthcare data support the development of strong synergies between research teams using these methods as part of a multi-disciplinary approach, which requires a dialogue between quantitative and qualitative methodologists and health professionals. The Araqnée workshop represents a first step towards the development of such synergies.

Key benefits of SNA for health service research

Despite methodological challenges, the application of SNA to healthcare data enables us to address new research questions in the field of health service research. It complements common approaches such as mapping, or theoretical frameworks such as diffusion theories, through

descriptive, probabilistic or modelling methods. It also opens up research perspectives that have not been explored much to date, including the study of evolutions over time by understanding determinants and mechanisms of change within healthcare provider networks.

Furthermore, the application of SNA to healthcare data provides opportunities to support decision-making in a context where collaborations between health professionals are at the centre of coordination issues, with implications for quality of care and new funding mechanisms. While the direct translation of research projects on healthcare provider networks into evidence-based decision-making has been minimal, due to the rather recent development of such projects, the networks identified by Thérèse Stukel and her colleagues (Stukel *et al.*, 2013) have inspired the creation of integrated systems to improve care for high-need, high-cost patients in Ontario. The identification of pre-existing informal healthcare provider networks can indeed form a rational basis for developing more formal networks, such as ACOs in the US, or to monitor care performance without limiting the attribution of responsibility to a single provider. However, such a transfer from research to decision-making requires further work on the best ways that health authorities and local organizations might use the tools developed by researchers to improve coordination. In addition, researchers should ensure that their methods are clear as SNA, especially when it uses modelling tools such as exponential random graph models, can be complex. The use of visualisation tools can be an important way of communicating findings to policy makers but their scientific soundness (in particular transparency regarding the graphics layout algorithms used) should be verified.

Future developments

In order to identify additional research projects applying SNA to healthcare data and to support peer-reviewed publications in this field, a call for papers will be announced for a special issue of the Francophone Journal on Health and Territories (*Revue francophone sur la santé et les*

territoires, RFST)² focusing on "Network analysis and coordination in health".

Acknowledgements

We are grateful to all the institutions that provided funding or support for the organization of this workshop (Irdes, Iresp, GDR 3771, EHESP, Cist), and to the workshop's speakers and attendees for their input.

The author also thanks the scientific committee of the Araquée workshop (Laurent Beauguitte, Magali Coldefy, Charène Le Neindre, Nolwenn Le Meur and Thérèse Stukel) who provided constructive feedback to improve this article.

Conflict of interest

The author declares no conflict of interest.

References

- Bach, P.B., Pham, H.H., Schrag, D., Tate, R.C. & Hargraves, J.L. (2004). Primary care physicians who treat blacks and whites. *New English Journal of Medecine*, 351(6), 575–584.
- Barnett, M.L., Landon, B.E., O'Malley, A.J., Keating, N.L. & Christakis, N.A. (2011). Mapping physician networks with self-reported and administrative data. *Health services research*. 46(5), 1592–1609.
- Bergeron, H. & Castel, P. (2010). Captation, appariement, réseau : une logique professionnelle d'organisation des soins. *Sociologie du travail*. 52(4), 441–460.
- Cunningham, F.C., Ranmuthugala, G., Plumb, J., Georgiou, A., Westbrook, J.I. & Braithwaite, J. (2012). Health professional networks as a vector for improving healthcare quality and safety: a systematic review. *BMJ Qual. Saf.* 21(3), 239–249.
- Gandr , C. (2018). Analyse de r seau pour l' tude des r seaux de professionnel.le.s et d' tablissements de sant  (Sunbelt 2018). ARCS - *Analyse de r seaux pour les sciences sociales*.
- Landon, B.E., Keating, N.L., Barnett, M.L., Onnela, J.-P., Paul, S., O'Malley, A.J., Keegan, T. & Christakis, N.A. (2012). Variation in patient-sharing networks of physicians across the United States. *JAMA*, 308(3), 265–273.

² <https://rfst.hypotheses.org/>

- Landon, B.E., Keating, N.L., Onnela, J.-P., Zaslavsky, A.M., Christakis, N.A. & O'Malley, A.J. (2018). Patient-sharing networks of physicians and health care utilization and spending among Medicare beneficiaries. *JAMA internal medicine*, 178(1), 66–73.
- Landon, B.E., Onnela, J.-P., Keating, N.L., Barnett, M.L., Paul, S., O'Malley, A.J., Keegan, T. & Christakis, N.A. (2013). Using administrative data to identify naturally occurring networks of physicians. *Medical care*, 51(8), 715–721.
- Lazega, E. (2001). *The collegial phenomenon: the social mechanisms of cooperation among peers in a corporate law partnership*. Oxford: Oxford University Press.
- Le Meur, N., Ferrat, L., Gao, F., Quidu, F. & Louazel, M. (2017). Maillage territorial des établissements de santé : apport des modèles issus de la théorie des graphes. *Journée de gestion et d'économies médicales*, 35(4), 197–208.
- Lefebvre, B. (2013). Les chaînes hospitalières en Inde. Quels modèles pour approcher la diffusion spatiale de ces réseaux de soins ?, *Actes de la deuxième journée d'étude du groupe fmr*.
- Nekkab, N., Astagneau, P., Temime, L. & Crépey, P. (2017). Spread of hospital-acquired infections: A comparison of healthcare networks. *PLoS computational biology*, 13(8), e1005666.
- Pinell, P. (2005). Champ médical et processus de spécialisation. *Actes de la recherche en sciences sociales*, 1, 4-36.
- Shelton, R.C., Lee, M., Brotzman, L.E., Crookes, D.M., Jandorf, L., Erwin, D. & Gage-Bouchard, E.A. (2019). Use of social network analysis in the development, dissemination, implementation, and sustainability of health behavior interventions for adults: A systematic review. *Social science & medicine*, 220, 81–101.
- Stukel, T.A., Glazier, R.H., Schultz, S.E., Guan, J., Zagorski, B.M., Gozdyra, P. & Henry, D.A. (2013). Multispecialty physician networks in Ontario. *Open medicine*, e40-55.