MOX-Report No. 07/2018

Network Analysis of Comorbidity Patterns in Heart Failure Patients using Administrative Data

Ieva, F.; Bitonti, D.

NETWORK ANALYSIS OF COMORBIDITY PATTERNS IN HEART FAILURE PATIENTS USING ADMINISTRATIVE DATA

Francesca Ieva* and Daniele Bitonti*

*MOX– Modeling and Scientific Computing
Department of Mathematics
Politecnico di Milano
Via Bonardi 9, 20133 Milano, Italy

Abstract

Background: Congestive Heart Failure (HF) is a widespread chronic disease characterized by a very high incidence in elder people. The high mortality and readmission rate of HF strongly depends on the complicated morbidity scenario often characterising it.
Methods: Data were retrieved from the healthcare administrative datawarehouse of Lombardy, the most populated regional district in Italy. Network analysis techniques and community detection algorithms are applied to comorbidities registered in hospital discharge papers of HF patients, in 7 cohorts between 2006 and 2012.
Results: The relevance network indexes applied to the 7 cohorts identified death, ipertension, arrythmia, renal and pulmonary diseases as the most relevant nodes related to HF, in terms of prevalence and closeness/strenght of the relationship. Moreover, 3 clusters of nodes have been identified in all the cohorts, i.e. those related to cancer, lung diseases and heart/circulation related problems.
Conclusions: Network analysis can be a useful tool in epidemiologic framework when relational data are the objective of the investigation, since it allows to visualize and make inference on patterns of association among nodes (here HF comorbidities) by means of both qualitative indexes and clustering techniques.

Keywords: Network Analysis; Administrative databases; Heart Failure; Comorbidities.

Introduction

Congestive Heart Failure (HF in the following) is a widespread chronic disease characterized by a very high incidence in elder people [**1**]. HF prevalence steeply increases with aging [**2**]. One year mortality ranges from 35-40% and more than 50% of patients are readmitted to hospital between 6 months and 1 year after the diagnosis, due to a complicated morbidity scenario, among others. In this epidemiological setting, elders with HF are representative of a growing segment living longer with chronic conditions prone to multiple transitions from hospital to home and vice versa. This unavoidably affects their quality of life, and turns in an important healthcare management and costs issue. Last but not least, in such a context it is pretty unreasonable to consider the health status of a patient as due to a “main” disease surrounded by other possible minor diseases. It is more often the case that more than one condition contributes to determine the health need and consumption.

Another issue related to HF and related healthcare practice and management is the following: it is more and more common nowadays to make use of secondary databases to conduct epidemiological enquires concerning HF. In fact, patients with HF randomized in controlled trials are generally selected and do not fully represent the “real world” [**3**].

For all these reasons, the objective of our study (more details in [**4**]) is to show the potential, the usefulness and the advantages of applying Network analysis ([**5**], [**6**], [**7**]) and in general a relational approach in the study of the comorbidities recorded in hospitalizations charts of HF patients [**8**]. Specifically, we wish to investigate relationships among the possible evolution of morbidities patterns accompanying HF and their relationship with death over the years. targeting this goal for the first time in literature using administrative data ([**9**],[**10**]). This allows us to study the pathologies taking advantage of techniques of network analysis.

The article is organized as follow: after an introduction to the basics of network analysis and a brief description of data, we illustrate the applications of network analysis to our data and finally the results’ discussion.

Methods

Network analysis in a nutshell

A network is a graph with N nodes (or vertices) and L links (or edges) that can be weighted or unweighted, directed or not. An unweighted network is completely represented by its $N \times N$ adjacency matrix A such that $A_{ij} = 1$ if node i points to node j, $A_{ij} = 0$ otherwise.
Let $G = (V; E)$ be a graph, where V is the set of its vertices such that $|V| = N$ and E is the set of its edges such that $|E| = L$. There are many important properties through which a network can be described ([**5**],[**7**]). In [**1**] and references therein, all the definition needed for the following analyses both in terms of centrality/relevance indexes of nodes as well as communities can be found.

Setting

Data were retrieved from the healthcare administrative datawarehouse of Lombardy, a region of Italy which accounts for about 16% (almost ten million) of its population. The

National Health Service covers the entire Italian population and in Lombardy this has been associated since 1997 with an automated system of databases to collect a variety of information. Full details about the dataset and selection criteria of the cohort are reported in [**11**].

Data analysis

Analyses are carried out with R software ([**12**],[**13**]) and network dedicated packages, like `igraph` [**14**].

We consider only the last hospitalization of each patient in the period 2006-2012, since it is assumed to describe his/her most compromised clinical condition. In doing so, we end up with 7 networks (i.e., 7 cohorts), obtained by projection of the bipartite network "patients-comorbidity" on the "comorbidity" dimension. Each patients contributes to the network related to the year of his/her last hospitalization. Nodes are represented by comorbidities (Death is a node of the comorbidity network, since we want to identify which pathologies are most connected to it). Two nodes are connected by an edge, weighted according to the amount of patients presenting that couple of comorbidities. The strength of the association between two nodes is measured in terms of ϕ-correlation [**15**]. For each patient, in addiction to the comorbidities and death/survival indicators, information about age [years] and gender are available.

From the procedure described above, we get a dense network [**5**], which is odd to treat both from a modelling and computational point of view. Therefore, a thresholding [**6**] is needed, and we adopted the following criterion: let G be the undirected network under study, and τ a prescribed or desired density for the network. Then the network density (defined as ρ = L/[N(N-1)/2], where L and N the number of links and nodes of the network G, respectively) can be tuned in order to maintain edges only if they fulfill the requirement $\phi > \tau$.

Figure 1 shows networks concerning the years 2006 and 2012, with $\tau = 0.02$. The shape of the nodes are defined according to the presence of men (higher if the node is square shaped) or women (higher if the node is circle shaped) presenting that pathology, and the colors are related to the prevalence.

For each node in each network, an index of relevance is computed. The index is composed by *degree centrality* (measuring the number of connected pathologies), *strenght* (measuring the strength of the connection with other patologies), *weighted local transitivity* or *closeness centrality* (measuring the proximity to other pathologies) and *prevalence* of that node. This allows to identify which nodes are more relevant within each network and within each year. Finally, a community detection algorithm based on modularity maximization ([**16**],[**17**],[**18**]) is applied in order to find relevant communities of nodes within the networks.

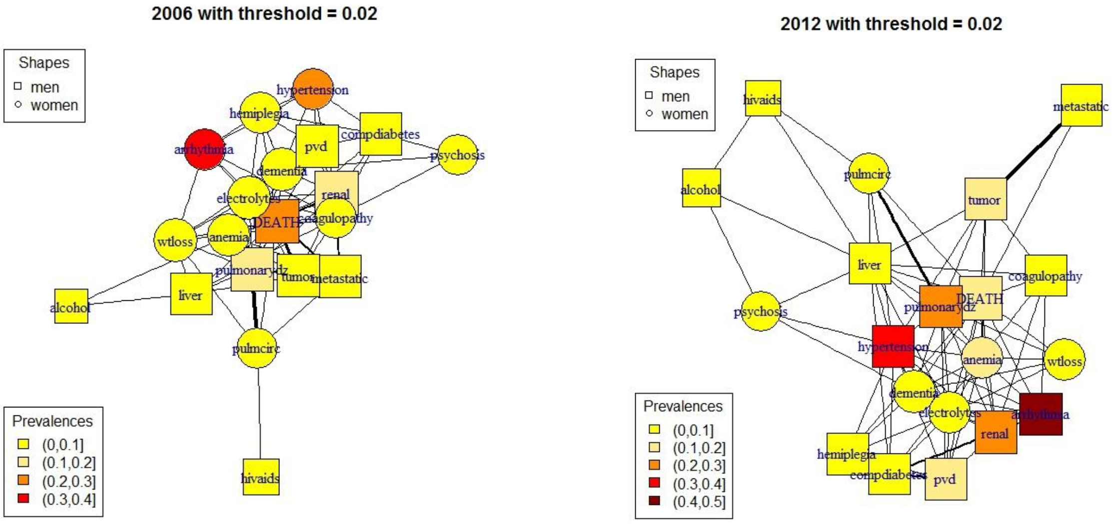

Figure 1: Representations of the 2006 (left panel) and 2012 (right panel) networks.

Results

The procedure described in the last Section results in 7 networks to be analyzed. We reduced the density of the graphs considering only links that had a ϕ-correlation greater than $\tau = 0.02$ (Figure 1).

The relevance indexes identified death, hypertension, arrhythmia, renal and pulmonary diseases as the most relevant nodes related to death, in terms of prevalence and closeness/strenght of the relationship.

Figure 2 shows the communities detected in 2007 and 2009 cohorts, which are those related to cancer, lung diseases and heart/circulation related problems. These communities are present in almost all the cohorts in the same configuration.

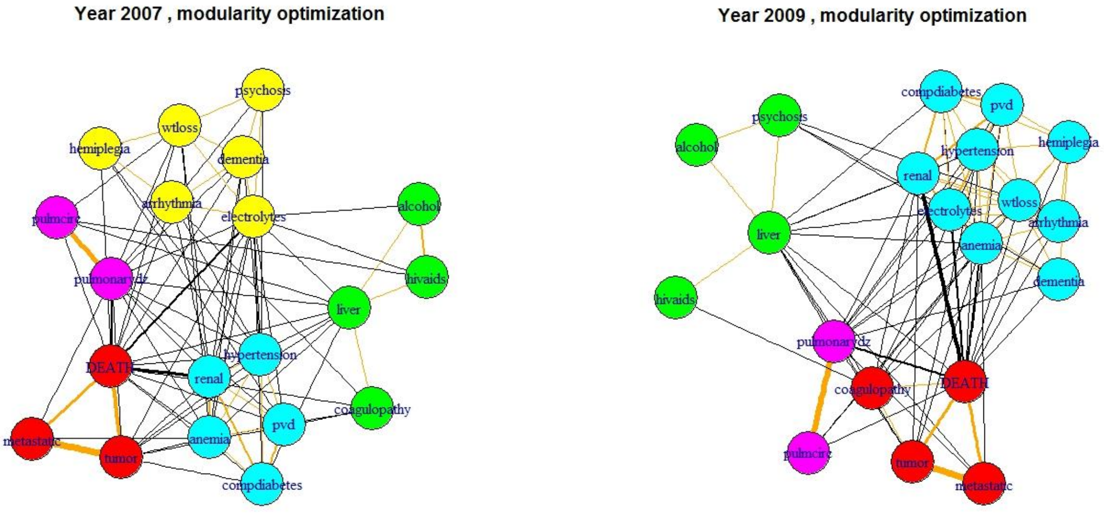

Figure 2: Communities of nodes (i.e., comorbidities) detected
in the 2007 (left panel) and 2009 (right panel) networks.

These results show that even in a simple example like the one proposed, patterns of connections among comorbidities related to HF may be discovered and monitored in their relationships with death. From these preliminary results, it seems that such patterns do not evolve along time. Further investigations are needed to consider potential risk profiles of patients to be monitored in dedicated programs.

Discussion and Further Developments

In this work we showed a promising approach to the analysis of comorbidity patterns in patients affected by HF using networks. It represents an innovative and flexible method that can be adopted for many different kind of epidemiological investigations. Many features that emerged thanks to the network approach we adopted might be exploited further through more classical statistical methods.

In general, network analysis can be considered a useful tool in epidemiologic framework when relational data are the objective of the investigation, since it allows to visualize and make inference on patterns of association among nodes (here HF comorbidities) by means of both qualitative indexes and clustering techniques. This is particularly relevant when the size of the network (i.e., the number o nodes) becomes high.

Future developments of the present work may regard:

- to increase the size of the network, using DRGs instead of comorbidities;
- to consider bipartite networks of patients and comorbidities (or diagnoses) directly, without projecting and thresholding;
- to define an unique index that takes the prevalence, degree, strength and closeness into account, properly weighting their contributes (possibly according to clinicians' suggestions);
- to refine the community detection, exploiting techniques like stochastic block models (SBM) [**19**] or latent class models for bipartite newtorks.

Acknowledgements

The work has been developed within the HEAD project “Public Management Research: Health and Education Systems Assessment”, funded by Politecnico di Milano. The authors acknowledge the Project “Utilization of Regional Health Service databases for evaluating epidemiology, short- and medium-term outcome, and process indexes in patients hospitalized for heart failure”, for providing data.

References

[1] Bui, A.L., Horwich, T.B., Fonarow, G.C. (2011) Epidemiology and risk profile of heart failure. Nature Reviews Cardiology, **8**: 30-41.
[2] Curtis, L.H., Whellan, D.J., Hammill, B.G., Hernandez, A.F., Anstrom, K.J., Shea, A.M., Schulman, K.A. (2008) Incidence and prevalence of heart failure in elderly persons, 1994-2003. Archives of Internal Medicine, **168**(4): 418-424.
[3] Maggioni, A.P., Orso, F., Calabria, S., Rossi, E., Cinconze, E., Baldasseroni, S., Martini, N. (2016) The real-world evidence of heart failure: findings from 41 413 patients of the ARNO database. European Journal of Heart Failure, **18**(4): 402-10.
[4] Bitonti, D. (2016) Network Analysis of Comorbidity Patterns in Heart Failure Patients using Administrative Data. MD thesis, Politecnico di Milano LINK
[5] Barabasi, A.L. (2016) Network Science. Cambridge University Press; first edition, August 5, 2016.
[6] Kolaczyk, E.D. (2009) Statistical Analysis of Network Data - Methods and Models. Springer, New York, 2009.
[7] Newman, M.E.J. (2010) Networks - An Introduction. Oxford University Press, New York, 2010.
[8] Van Deursen, V.M., Urso, R., Laroche, C., Damman, K., Dahlstrom, U., Tavazzi, L., Maggioni, A.P., Voors, A.A. (2014) Co-morbidities in heart failure. Heart Failure Review, **19**(2): 163-72.
[9] Hoover, K.W., Tao, G., Kent, C.K., Aral, S.O. (2011) Epidemiologic research using administrative databases: garbage in, garbage out. Obstetrics and Gynecology, **117**(3): 729; author reply 729-30.
[10] Schneeweiss, S., Avorn, J. (2005) A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology, **58**(4): 323-37.
[11] Mazzali, C., Paganoni, A.M., Ieva, F., Masella, C., Scalvini, S., Frigerio, M. et al. (2016) Methodological issues on the use of administrative data in healthcare research: the case of heart failure hospitalizations in Lombardy Region, 2000 to 2012. BMC Health Services Research, **16**(1): 234 doi: 10.1186/s12913-016-1489-0
[12] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (2015). URL http://www.R-project.org/.
[13] Kolaczyk, E.D., Csardi, G (2014) Statistical Analysis of Network Data with R. Springer, New York, 2014.
[14] Csardi, G., Nepusz, T. (2006) The igraph software package for complex network research. Inter. Journal Complex Systems RIVEDERE REF.
[15] Hidalgo, C.A., Blumm, N., Barabasi, A.L., Christakis, N.A. (2009) A Dynamic Network Approach for the Study of Human Phenotypes. PLoS Computational Biology, **5**(4): e1000353.
[16] Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., Wagner, D. (2008) On Modularity Clustering. IEEE Transactions on Knowledge and Data Engineering, **20**(2):172-188.
[17] Newman, M.E.J. (2004) Fast algorithm for detecting community structure in networks. Physical review, E69(6), 066133.
[18] Girvan, M., Newman, M.E.J. (2002) Community structure in social and biological networks. Proceedings of the national academy of sciences, **99**(12): 7821-7826.
[19] Holland, P.W., Blackmond Laskey, K., Leinhardt, S. (1983) Stochastic blockmodels: First steps. Social Networks, **5**: 109-137.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

06/2018 Antonietti, P.F.; Mazzieri, I.
High-order Discontinuous Galerkin methods for the elastodynamics equation on polygonal and polyhedral meshes

05/2018 Pagani, S.; Manzoni, A.; Quarteroni, A.
Numerical approximation of parametrized problems in cardiac electrophysiology by a local reduced basis method

03/2018 Antonietti, P. F.; Houston, P.; Pennesi, G.
Fast numerical integration on polytopic meshes with applications to discontinuous Galerkin finite element methods

04/2018 Ekin, T.; Ieva, F.; Ruggeri, F.; Soyer, R.
Statistical Medical Fraud Assessment: Exposition to an Emerging Field

01/2018 Berrone, S.; Bonito, A.; Stevenson, R.; Verani, M.
An optimal adaptive Fictitious Domain Method

02/2018 Canuto, C.; Nochetto, R. H.; Stevenson, R.; Verani, M.
A saturation property for the spectral-Galerkin approximation of a Dirichlet problem in a square

68/2017 Paolucci, R.; Infantino, M.; Mazzieri, I.; Özcebe, A.G.; Smerzini, C.; Stupazzini, M.
3D physics-based numerical simulations: advantages and current limitations of a new frontier to earthquake ground motion prediction. The Istanbul case study.

67/2017 Esterhazy, S.; Schneider, F.; Mazzieri, I; Bokelmann, G.
Insights into the modeling of seismic waves for the detection of underground cavities

66/2017 Tamellini, M.; Parolini, N.; Verani, M.
An optimal control problem of two-phase compressible-incompressible flows

64/2017 Fumagalli, I.; Parolini, N.; Verani, M.
Optimal control in ink-jet printing via instantaneous control