



AFRL-AFOSR-UK-TR-2022-0064

Path weights in undirected Markov random fields

Roverato, Alberto
ALMA MATER STUDIORUM UNIVERSITA DI BOLOGNA SEDE
VIA ZAMBONI 33
BOLOGNA, , 40126
IT

06/06/2022
Final Technical Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20220606	2. REPORT TYPE Final	3. DATES COVERED	
		START DATE 20161115	END DATE 20211114
4. TITLE AND SUBTITLE Path weights in undirected Markov random fields			
5a. CONTRACT NUMBER	5b. GRANT NUMBER FA9550-17-1-0039	5c. PROGRAM ELEMENT NUMBER	
5d. PROJECT NUMBER	5e. TASK NUMBER	5f. WORK UNIT NUMBER	
6. AUTHOR(S) Alberto Roverato			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ALMA MATER STUDIORUM UNIVERSITA DI BOLOGNA SEDE VIA ZAMBONI 33 BOLOGNA 40126 IT			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) EOARD UNIT 4515 APO AE 09421-4515		10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/AFOSR IOE	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-AFOSR-UK-TR-2022-0064
12. DISTRIBUTION/AVAILABILITY STATEMENT A Distribution Unlimited: PB Public Release			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT This research work has focused on statistical inference for network models. Following the direction given in the original submission, the team developed a methodology for paths and path weights. The main aim was to clarify the role played by the path weight in undirected graphical models providing a clear interpretation to the value they take. This goal was successfully achieved and has lead to 5 articles published in top statistical journals. These results were then extended to be made applicable to the analysis of symmetric structures in brain networks form from functional magnetic resonance imaging and the computation of centrality measures in food networks encoding food consumption patterns. In order to increase the dissemination of the research results, Dr. Roverato wrote a book "Graphical Models for Categorical Data", published by Cambridge University Press. Moreover, the research developed under this grant was presented at 14 conferences and professional meetings. Finally, Dr. Roverator co-authored the chapter of a book published by Chapman and Hall "Handbook of Graphical Models".			
15. SUBJECT TERMS			
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	SAR 5
19a. NAME OF RESPONSIBLE PERSON MARK FRIEND			19b. PHONE NUMBER (Include area code) 314-235-6292

Final Report for EOARD Grant FA9550-17-1-0039

“Path weights in undirected Markov random fields”

January 31st 2022

Name of Principal Investigator

- **e-mail address:** alberto.roverato@unipd.it
- **Istitution:** Department of Statistical Sciences, University of Padova, Italy.
- **Mailing Address:** Via Cesare Battisti, 241 - 35121 Padova, Italy.
- **Phone:** +39 0498274183
- **Fax:** +39 0498274170

Period of performance: 11/15/2016 - 11/14/2021

Abstract

Graphical models are an elegant framework that combines uncertainty and graph theory to represent complex phenomena. A graph, or network, is a structure consisting of a set of objects, called vertices, and a set of connections between pairs of vertices, called edges. The vertices of the network associated with a graphical model are the variables of the model and the edges describe how the variables interact with each other. The edges missing from the network can be interpreted as absence of interaction. An appealing characteristic of networks is that they can be represented graphically and many of their features and properties can be understood from the visual inspection of their graphical representation. The network greatly simplifies the interpretation of the model, making its independence structure more immediate and intuitive. This also facilitates the communication of the scientific contents of the model to researchers who are not familiar with the statistical formalism.

A path in a network is an alternating sequence of vertices and edges which begins with a vertex and ends with a different vertex. A pair of vertices may be connected by an edge, representing a direct association, but also by a number of different paths representing indirect associations mediated by the intermediate variables of the path. In graphical models, paths joining vertices are the main tools used in the specification of the association structure of the variables. In models where edges are represented by directed arrows it is possible to compute path coefficients. Specifically, the theory of path analysis provides a method that aims at quantifying the relative importance of causal relationships represented by directed paths. On the other hand, in models for undirected networks, where edges are represented by undirected lines, little has been investigated on the strength of the association encoded by paths.

The research work concerning this project has focused on statistical inference for network models. Following the direction given in the original submission, we have developed of a methodology for paths and path weights. The the main aim was to clarify the role played by the path weight in undirected graphical models providing a clear interpretation to the value they take. This result has been successfully achieved and has lead to articles published in top statistical journals such as Roverato and Castelo (2017), Roverato and Castelo (2018) and Roverato and Castelo (2020). These results were then extended so as to be applied in a range of different contexts. More specifically, we considered the analysis of symmetric structures in brain networks form from functional magnetic resonance imaging (fMRI) data (Ranciati et al., 2021) and the computation of centrality measures in food networks encoding food consumption patterns (Roverato, 2021b). In order to increase the dissemination of the research results, I have written a book, published by Cambridge University Press, (Roverato, 2017) and co-authored the chapter of a book published by Chapman and Hall/CRC (La Rocca and Roverato, 2019). Furthermore, I have contributed to the discussion of scientific papers on biological networks by other authors (Roverato, 2021a).

List of Publications

La Rocca, L. and A. Roverato (2019). Discrete graphical models. In M. Drton, S. L. Lauritzen, M. Maathuis, and M. Wainwright (Eds.), *Handbook of Graphical Models*, Handbooks of Modern Statistical Methods. Chapman and Hall/CRC.

Ranciati, S., A. Roverato, and A. Luati (2021). Fused graphical lasso for brain networks with symmetries. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 70(5), 1299–1322.

Roverato, A. (2017). *Graphical Models for Categorical Data*. SemStat Elements. Cambridge University Press.

- Roverato, A. (2021a). Discussion to: Bayesian graphical models for modern biological applications by Y. Ni, V. Baladandayuthapani, M. Vannucci and F.C. Stingo. *Statistical Methods & Applications*.
- Roverato, A. (2021b). On the interpretation of inflated correlation path weights in concentration graphs. *Statistical Methods & Applications* 30(5), 1485–1505.
- Roverato, A. and R. Castelo (2017). The networked partial correlation and its application to the analysis of genetic interactions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 66(3), 647–665.
- Roverato, A. and R. Castelo (2018). Differential networking with path weights in gaussian trees. In V. Kratochvíl and M. Studený (Eds.), *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, Volume 72 of *Proceedings of Machine Learning Research*, Prague, Czech Republic, pp. 404–415. PMLR.
- Roverato, A. and R. Castelo (2020). Path weights in concentration graphs. *Biometrika* 107(3), 705–722.

Significant collaborations that resulted from the project

Work concerning the research project has been carried out in collaboration with Prof. Robert Castelo (Pompeu Fabra University, Barcelona, Spain) and with this aim I visited Pompeu Fabra University regularly during the grant period, more specifically I spent visiting periods in February 2017, February 2018, August 2018, August 2019 and December 2019.

I also spent a longer visiting period at the Department of Mathematics of the University of Canterbury in Christchurch, New Zealand from June 15th to August 31st 2017. This visiting fellowship, which was partially co-financed by the University of Canterbury through an Erskine grant, aimed at two main goals. Firstly, to establish a collaboration with the Biomathematics Research Centre (within the Department of Mathematics and Statistics) thereby extending the application of the methodology of path weights to gene expression data with special attention to the problem of " $p \gg n$ ", i.e. when the number of variables exceeds the sample size. Secondly, to investigate, in collaboration with Prof. Marco Reale, the potential applications of the methodology of path weights in the area of finance as well as to economic (time series) data.

Finally, I was a visiting fellow from June 24th to July 5th at the Department of Mathematics of Brunel University in London to interact with Prof. Veronica Vinciotti on some specific aspects concerning statistical inference methods for brain networks.

The project has also involved young researchers at the beginning of their academic life. More specifically, the budget justification of the grant also included the amount of 10.000 dollars for a 4 month position for a post-doc student. The Department of Statistical Sciences of the University of Bologna provided additional financial support and, consequently I was able issue a 15 months research contract to PhD Saverio Ranciati. The involvement of Dr. Ranciati in this project concerns the application of the methodological results of our research to the identification of brain networks from fMRI data. Furthermore, I had two PhD students collaborating to this project: PhD Violetta Zoffoly and PhD Dung Ngoc Nguyen.

The research results of this project were presented at a number of scientific conferences, detailed below.

- *9th International Conference of the ERCIM Working Group on Computational and Methodological Statistics*, Sevilla, Spain December 9-11, 2016.
- *Department of Mathematics and Statistics University of Canterbury*, Christchurch, New Zealand, July 2017.
- *Department of Mathematics and Statistics University of Canterbury*, Christchurch, New Zealand, August 2017.
- *Van Dantzig Seminar - Dutch nationwide series of lectures in statistics* University of Amsterdam, Holland, October 2017, invited presentation.
- *Department of Economics* Università Cattolica del Sacro Cuore, Milan, Italy, November, 2017.
- *10th International Conference of the ERCIM Working Group on Computational and Methodological Statistics*, London, UK, December 2017, invited presentation.
- *Workshop on Statistical Network Science*, Brunel University, London, UK, June 2018.
- *The 9th International Conference on Probabilistic Graphical Models*, Prague, Czech Republic, September 2018.
- *The 11th International Conference of the ERCIM Working Group on Computational and Methodological Statistics* Pisa, Italy, December 2018, invited presentation.
- *COSTNET 2018 conference - European Cooperation for Statistics of Network Data Science*, Warsaw, Poland, September 2018.
- *Data Science Seminars: Department of Economics and Business*, University Pompeu Fabra, Barcelona, Spain, March 2019.

- *COSTNET 2019 conference - European Cooperation for Statistics of Network Data Science*, Bilbao, Spain. October 2019, two distinct presentations, one invited and one contributed.
- *Joint Statistical Meetings* Denver, Colorado, USA, August 2019, invited presentation.
- *COSTNET 2020 conference - European Cooperation for Statistics of Network Data Science*, Virtual Conference Planned by the Department of Statistics at LMU Munich (Germany), September 2020, invited presentation.