

University of Groningen

Extensions of graphical models with applications in genetics and genomics

Behrouzi, Pariya

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Behrouzi, P. (2018). *Extensions of graphical models with applications in genetics and genomics*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Samenvatting

In deze these bespreken we enkele problemen gerelateerd aan het modelleren van complexe systemen. Velden zoals systeemgenetica, systeembioïogie, epidemiologie, en bioïnformatica gebruiken vaak grootschalige modellen, waarin duizenden componenten op complexe wijze met elkaar verbonden zijn. Mogelijk het meest onderscheidend van de grafische modelleerbenadering is de geschiktheid voor het formuleren van probabilistische modellen van complexe fenomenen in toegepaste velden, terwijl controle wordt behouden over de benodigde rekenkosten. In de echte wereld zijn niet alle datasets continu. Discrete data, of datasets met een combinatie van gemengde en continue data, zijn veelvoorkomend in de bovengenoemde velden.

In Hoofdstuk 2 introduceren we een methode voor reconstructie van een conditioneel onafhankelijkheidsnetwerk uit niet-Gaussische data. Deze methode is specifiek geschikt voor ordinale data en gemengd ordinale en continue data. Dit soort data komt veel voor in de systeemgenetica, waar de hoofdfocus is om beter begrip te ontwikkelen over de stroom van biologische informatie onderliggend aan complexe eigenschappen. In dit hoofdstuk richten we ons op de eigenschap “overleving”: we zoeken loci – locaties op een genoom – die niet onafhankelijk conditioneel op andere loci segregeren. De netwerkschatting berust op gepenaliseerde Gaussische copula grafische modellen. Dit neemt een groot aantal markers en een klein aantal individuen in beschouwing.

In Hoofdstuk 3 breiden we het sparse copula grafische model, zoals voorgesteld in Hoofdstuk 2, uit om hoogkwalitatieve linkage maps voor biparentale diploïde en polyplöide soorten te construeren. Een linkage map bevat genetische informatie, zoals het aantal chromosomen van een soort, het aantal markers in ieder chromosoom, en de volgorde van de markers in ieder chromosoom. In de voorgestelde methode van map constructie ontdekken we linkage groepen, meestal chromosomen, en de volgorde van de markers binnen iedere linkage groep. Dit doen we door de conditionele onafhankelijkheidseigenschappen tussen grote aantallen markers in ieder chromosoom af te leiden uit genotyping studies, zoals genome-wide association studies.

In Hoofdstuk 4 introduceren we een R package netgwas om de methodes uit hoofdstukken

2 en 3 efficient toe te passen. Dit package bevat een set hulpmiddelen, gebaseerd op ongerichte grafische modellen, om drie belangrijke en gerelateerde doeleindes in genetica en genomica te bereiken: constructie van linkage groepen, intra- en interchromosomale interacties, en hoogdimensionale genotype-fenotype (en genotype-fenotype-omgeving) interactienetwerken.

In Hoofdstuk 5 introduceren we sparse dynamic chain graph modellen voor netwerk inference in hoogdimensionale niet-Gaussische tijdseriesdata. De voorgestelde methode wordt geparametriseerd met behulp van een precision matrix, die intra time-slice conditionele onafhankelijkheden tussen variabelen op een vast tijdstip encodeert, en een autoregressieve coefficient, die dynamische conditionele onafhankelijkheidsinteracties tussen tijdseriecomponenten over opeenvolgende tijdstappen bevat. We passen onze methode toe op een Nederlandse Studie naar Depressie en Angst (NESDA) dataset om psychologische factoren te bepalen die invloed hebben op de ontwikkeling en langetermijnsprognose van angst en depressie.

Acknowledgements

First and foremost, I would like to express my gratitude to **Prof. Ernst Wit** who gave me the opportunity to work in his group and develop my skills in research. His guidance, support, and encouragement throughout all the different phases of this thesis were invaluable to the success of my PhD.

I would like to express my gratitude to **Prof. Ritsert Jansen** for supporting me and giving me the opportunity to work at the Groningen BioInformatics Centre (GBIC). I am also grateful to **Prof. Frank Johannes** for giving me the chance to join his group at GBIC. I appreciate his enthusiasm to let me work on my new statistical method on his data.

My sincere thanks goes to **Prof. Korbinian Strimmer**, **Prof. Christine Gräfin zu Eulenburg**, and **Prof. Edwin van den Heuvel** for assessing this thesis.

I gratefully acknowledge the funding received for a Short Term Scientific Mission from the **COSTNET Program**. Chapter 5 of my thesis is the outcome of this visit; my special thanks to **Dr. Fentaw Abegaz** for this fruitful collaboration and many interesting discussions.

I am grateful to my friends and former colleagues: **Javier**, **Antonio**, **Ivan**, **Lotsi**, **Spyros**, **Pancho**, and **Mehdi** for the stimulating discussions, and for all the fun we had in the last few years. I would also like to thank everyone at GBIC such as **René**, **Konrad**, **Lionel**, and **Maria** for all the interesting discussions we had during lunches and coffee breaks. I am thankful to **Danny** for helping me learn multi-core programming in R language and using the HPC cluster of the University of Groningen. Thanks to my new colleague **Guus** at the Wageningen University for translating my summery into Dutch.

A special thanks to **Nynke** for her friendship and support. You made my stay in the Netherlands very enjoyable. Thank you for being there for me whenever I need it. I would also like to thank **Vladi** for being a great friend, his support, and reading the last chapter of my thesis. Many thanks to **Frank** and **Esther**, **Mojde** and **Masoud**, **Samaneh**, **Azadeh** and **Soheil** for their friendship and support. In this journey I met new people in conferences in particular, **Maral** who became a great friend: thank you for your friendship and being my paranymph. Special thanks to **Anneke** and the late **Dick**, who made tolerable the difficulties of leaving our families behind.

This work is dedicated to **my loving parents**, who raised me with a love of science and gave me the unconditional love and endless support through all the years. I owe all my accomplishment to them. None of this would have been possible without the support and encouragement of my family. I am thankful to my older sisters **Roya** and **Shiva** and my younger brothers **Amirreza** and **Hamidreza** for their love; I just could not imagine my life without them. My brothers-in-law **Alireza** and **Peyman** I can't thank them enough for their kindness, caring, and being a great friend.

My deepest gratitude goes to **Reza** who stood by me all the time throughout my research work and my difficult times. We built our lives together in the Netherlands and we shared every moment of it. Thank you for all your love, support, encouragement, and understanding during this time. I cannot find a better way of thanking you than to dedicate this thesis to you.

Pariya Behrouzi
Groningen
January, 2018

ISBN (Print): 978-94-034-0321-2
ISBN (Digital): 978-94-034-0320-5

Colophon

This thesis was completed using the PhD thesis L^AT_EX template, by Krishna Kumar,
University of Cambridge.

Printed by: Proefschriften (info@printservice-edu.nl)