



Universiteit
Leiden
The Netherlands

Bayes and Networks

Gao, F.

Citation

Gao, F. (2017, May 23). *Bayes and Networks*. Retrieved from <https://hdl.handle.net/1887/49012>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/49012>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/49012> holds various files of this Leiden University dissertation.

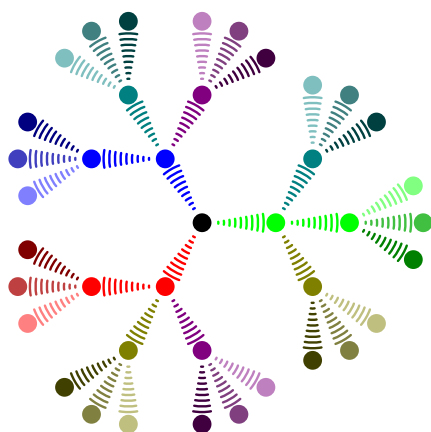
Author: Gao, F.

Title: Bayes and networks

Issue Date: 2017-05-23

FENGNAN GAO

¡ BAYES & NETWORKS



Shanghai, April 2017

FENGNAN GAO: *Bayes & Networks*, Dirichlet-Laplace Deconvolution and Statistical Inference in Preferential Attachment Networks, © April 2017

The author designed the cover by himself. The bottom-right corner lies a phoenix, which FENG in the author's name stands for in the Chinese language.

Title Page: The decoration on the margin was modified from the code published on \TeX StackExchange by Gonzalo Medina. The beautiful network illustration was distributed by Till Tantau, the author of *TikZ* under the GNU Free Documentation License.

All rights reserved. No part of this publication may be reproduced in any form or by any electronical or mechanical means including informaiton storage and retrieval systems without the prior written permission from the author.

A catalogue record is available from the Leiden University Library.

The research in the dissertation was supported by the Netherlands Organization for Scientific Research (NWO).



Netherlands Organisation
for Scientific Research



Universiteit
Leiden

Bayes & Networks

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 23 mei 2017
klokke 10:00 uur

door

FENGNAN GAO
geboren te Jiangsu, China
in 1988

Samenstelling van de promotiecommissie:

Promotor:

Prof. dr. A. W. van der Vaart (Universiteit Leiden)

Overige Leden:

Prof. dr. B. de Smit (Universiteit Leiden, voorzitter)

Prof. dr. J. J. Meulman (Universiteit Leiden, secretaris)

Prof. dr. R. W. van der Hofstad (TU Eindhoven)

Prof. dr. J. H. van Zanten (Universiteit van Amsterdam)

Dr. R. M. Castro (TU Eindhoven)

Dr. A. J. Schmidt-Hieber (Universiteit Leiden)

To my family

回首向來蕭瑟處
歸去
也無風雨也無晴

蘇軾

宋神宗元豐五年

Looking back over the bleak passage survived,
The return in time,
Shall not be affected by windswept rain or sunshine.

SU SHI (1082)

CONTENTS

I	NONPARAMETRIC BAYESIAN DIRICHLET-LAPLACE DE-CONVOLUTION	1
1	POSTERIOR CONTRACTION RATES FOR DECONVOLUTION OF DIRICHLET-LAPLACE MIXTURES	3
1.1	Introduction	3
1.2	Notation and Preliminaries	6
1.3	Main Results	7
1.4	Finite Approximation	8
1.5	Entropy	12
1.6	Prior Mass	15
1.7	Proof of Theorem 1.1	17
1.8	Proof of Theorem 1.2	20
1.9	Normal Mixtures	21
II	STATISTICAL INFERENCE IN PREFERENTIAL ATTACHMENT NETWORKS	23
2	INTRODUCTION TO NETWORKS	25
2.1	Network Science	25
2.1.1	The emergence of network science	25
2.1.2	Fundamentals of graph theory	29
2.1.3	Properties of typical networks	30
2.2	Preferential Attachment Networks	31
2.2.1	History and motivation of the PA networks	31
2.2.2	A rather general PA model	34
2.2.3	The linear PA models with random initial degrees	35
2.2.4	The general sublinear PA models	35
2.2.5	The general sublinear parametric PA models	35
3	ESTIMATION OF GENERAL PA NETWORKS	37
3.1	Introduction	37
3.2	Empirical Estimator	39
3.3	Branching Process	41
3.3.1	Rooted ordered tree	42
3.3.2	Branching process	42
3.3.3	The continuous random tree model	44
3.4	Consistency	46
3.5	Simulation Studies	49
3.5.1	Sample variance study	51
3.5.2	Asymptotic normality?	52

4	ESTIMATION OF AFFINE PA NETWORKS	57
4.1	Introduction and Notation	57
4.2	Construction of the MLE	61
4.3	Consistency	65
4.4	Asymptotic Normality	71
4.5	Local Asymptotic Normality and Efficiency	75
4.6	The Case of Fixed Initial Degree	76
4.7	Quasi-Maximum-Likelihood Estimator	76
4.8	Simulation Study	81
4.8.1	On the shoulder of the giants	82
4.8.2	The majority rules	85
5	ESTIMATION OF PARAMETRIC PA NETWORKS	87
5.1	Introduction and Notation	87
5.2	Construction of the MLE	88
5.3	Consistency	90
5.4	Asymptotic Normality	102
5.5	A Remedy to a Historical Problem	107
III	MODELING THE DYNAMICS OF THE MOVIE-ACTOR NETWORK	109
6	MODELING THE DYNAMICS OF THE MOVIE-ACTOR NETWORK OF THE INTERNET MOVIE DATABASE	111
6.1	Introduction	111
6.2	Conceptual Model Description	113
6.3	Empirical Fitting to the IMDb Dataset	115
6.3.1	Movie sizes	115
6.3.2	Number of new actors	117
6.3.3	PA function	117
6.3.4	PA function on movie degrees	123
6.3.5	Model fitting	124
6.4	Simulations	124
6.5	Theoretical Study	128
6.6	Conclusion and Future work	130
IV	APPENDIX	131
A	DIRICHLET PROCESSES AND CONTRACTION RATES RELATIVE TO NON-METRICS	133
A.1	Dirichlet Processes	133
A.2	Contraction Rates Relative to Non-metrics	133
B	CONVERGENCE TO A POWER LAW OF THE MOVIE DEGREES IN THE PAM-IMDB MODEL	135
B.1	Introduction and Heuristics	135
B.2	Proof of Theorem B.1	138

B.2.1	Concentration around the mean	139
B.2.2	Identification of the mean sequence	141

BIBLIOGRAPHY	157
---------------------	-----

SUMMARY	165
----------------	-----

SAMENVATTING	167
---------------------	-----

ACKNOWLEDGEMENTS	169
-------------------------	-----

CURRICULUM VITÆ	171
------------------------	-----

COLOPHON	173
-----------------	-----

LIST OF FIGURES

Figure 2.1	Seven Bridges of Königsberg	28
Figure 3.1	Boxplots of EE's in different settings.	50
Figure 3.2	Sample Variance Study of EE	52
Figure 3.3	QQ-Plots of Empirical Estimators	53
Figure 3.4	Histogram of Rescaled Empirical Estimator	54
Figure 3.5	Estimated Density of $\sqrt{n}(\hat{r}_2(n) - r_2)$ with different network sizes n	55
Figure 4.1	loglog Plot of Empirical Degree Distribution vs. Degree in PA Networks	83
Figure 4.2	Histogram of MLE in Affine PA networks	84
Figure 6.1	Histogram of movie sizes in 1947	115
Figure 6.2	loglog-Histogram of Movie Sizes	116
Figure 6.3	loglog-histogram of all movie sizes until the end of 2007	116
Figure 6.4	Ratio of New Actors of Movies in 1971	118
Figure 6.6	Actor Degree Evolution	120
Figure 6.7	Fitting a straight line on the loglog-Histogram on Actor Degrees	120
Figure 6.8	$\log(1 - \hat{F}_N(k))$ -vs.- $\log k$ Plot of Actor Degrees	121
Figure 6.9	Movie Degree Evolution	122
Figure 6.10	Fitting a straight line on the loglog-histogram starting from $k = 40$ on movie degrees	122
Figure 6.11	$\log(1 - \hat{F}_N(k))$ vs. $\log k$ plot of movie degrees in year 1947	123
Figure 6.12	Histogram of Movie Degrees by 1950	125
Figure 6.13	Movie Degree Comparisons Between Simulation and Real Dataset	126
Figure 6.14	Actor Degree Comparison between Simulation and Real Dataset	127

LIST OF TABLES

Table 2.1	Representative Networks	26
Table 4.1	Summary of the Performance of the MLE in Affine PA Networks	84