

Contents

Contents	vii
Thanks	xi
1 Introduction	1
1.1 History and Appetizer	1
1.2 Task and Work flow	3
1.3 Structure of the Thesis	3
2 Biological and Technical Background	7
2.1 Genes and DNA	7
2.2 Information Flow in Biological Cells	9
2.3 Two-Channel DNA Microarrays	14
2.3.1 Multi- and Time Course Experiments	16
2.4 Error Sources during Microarray Experiments	17
2.5 Magnaporthe grisea	19
3 Preprocessing of Microarray Measurement Data	23
3.1 Microarray Design of Experiments	23
3.2 Data Preprocessing and Normalization	27
3.2.1 Microarray Background Subtraction	28
3.2.2 Between Chip Brightness Normalization	29
3.2.3 Within Chip Normalization by Lowess	31
3.3 Statistical Analysis and Design of Experiments (cont.)	34
3.3.1 Fisher-Pitman-Test	35
3.3.2 Minimal Sample Sizes	36
3.4 Results	37
4 Interpolation of Time Course Data	41
4.1 Accuracy Estimation by Bootstrapping	42
4.1.1 Exact Bootstrapping	43
4.2 Smoothing Splines	48
4.3 Results	57
	vii

CONTENTS

5	Clustering of Time Course Data	59
5.1	A Short Introduction to Clustering	60
5.2	Data Standardization	61
5.3	Distance Measures	63
5.3.1	p-Minkowski Metric	63
5.3.2	Pearson's Product-Momentum Correlation Coefficient	64
5.3.3	L^p -Metric	65
5.4	Linkage Methods	66
5.4.1	Complete and Single Linkage	67
5.4.2	Average and Centroid Linkage	67
5.5	Clustering Methods	69
5.5.1	K-Means	69
5.5.2	Hierarchical Clustering	71
5.5.3	Density-Based Spatial Clustering of Applications with Noise	72
5.6	Quality Measures and Cluster Validation	77
5.6.1	Separation Indices	78
5.6.2	Robustness Indices	79
5.7	Time Course of Clusters	80
5.7.1	Cluster Medoid	80
5.7.2	Cluster Centroid	80
5.7.3	Cluster Smoothing Spline	81
5.8	Results	81
6	Calculation of a Gene Interaction Network	85
6.1	Discrete Linear Time-Invariant State Space Model	86
6.2	Results	89
6.2.1	Autonomy	89
6.2.2	Additional Dimension Reduction	91
6.2.3	Robustness with Respect to the Step Size	92
6.2.4	System Stability	92
7	Summary and Outlook	93
7.1	Design of Experiments	93
7.2	Data Processing	94
7.3	Estimation of Gene Expression Time Courses	95
7.4	Data Clustering and Modeling of the Interaction Network	96
7.5	Epilogue	98
A	Appendix: MA-Plots	99

CONTENTS

B	Appendix: An Algorithm for the Fisher-Pitman-Test	107
C	Appendix: Validation of Clusterings using the Spline Distance	111
D	Appendix: Clustering Results	113
E	Appendix: Approximation of the Clusters	125
	List of Figures	137
	List of Tables	141
	Notations	143
	Bibliography	147
	Index	162