

# Contents

|          |                                                            |           |
|----------|------------------------------------------------------------|-----------|
| <b>1</b> | <b>Overview of Network Inference</b>                       | <b>1</b>  |
| 1.1      | Inferring Causality                                        | 1         |
| 1.1.1    | Basics of Genomics                                         | 2         |
| 1.1.2    | Observation, Intervention, and Inference                   | 3         |
| 1.1.3    | The Data                                                   | 4         |
| 1.1.4    | Data Sources                                               | 4         |
| 1.1.5    | Basic Data Metrics                                         | 5         |
| 1.1.6    | How to Read This Book                                      | 8         |
| 1.1.7    | How to Use the Software                                    | 8         |
| <b>2</b> | <b>Step 1: Clustering Data</b>                             | <b>11</b> |
| 2.1      | Introduction                                               | 11        |
| 2.1.1    | Clustering                                                 | 11        |
| 2.1.2    | Biclustering                                               | 12        |
| 2.2      | cMonkey                                                    | 12        |
| 2.2.1    | What it Does                                               | 12        |
| 2.2.2    | The Data                                                   | 12        |
| 2.2.3    | The Strategy                                               | 12        |
| 2.2.4    | Walkthrough Example on Toy Data                            | 16        |
| 2.3      | Factor Analysis for Bicluster Acquisition (FABIA)          | 18        |
| 2.3.1    | What it Does                                               | 18        |
| 2.3.2    | The Data                                                   | 19        |
| 2.3.3    | The Strategy                                               | 19        |
| 2.3.4    | Walkthrough Example on Toy Data                            | 21        |
| <b>3</b> | <b>Step 2: Use Steady State Data for Network Inference</b> | <b>23</b> |
| 3.1      | Introduction                                               | 23        |
| 3.2      | Median-Corrected Z-Scores (MCZ)                            | 23        |
| 3.2.1    | What it Does                                               | 23        |
| 3.2.2    | The Data                                                   | 24        |
| 3.2.3    | The Strategy                                               | 24        |

|          |                                                        |           |
|----------|--------------------------------------------------------|-----------|
| 3.2.4    | Performance on Examples                                | 25        |
| 3.3      | Network Identification by Multiple Regression (NIR)    | 28        |
| 3.3.1    | What it Does                                           | 28        |
| 3.3.2    | The Data                                               | 28        |
| 3.3.3    | The Strategy                                           | 29        |
| 3.3.4    | Performance on Examples                                | 30        |
| 3.4      | Gene Network Inference with Ensemble of trees (GENIE3) | 34        |
| 3.4.1    | What it Does                                           | 34        |
| 3.4.2    | The Data                                               | 34        |
| 3.4.3    | The Strategy                                           | 34        |
| 3.4.4    | Performance on Examples                                | 37        |
| 3.5      | Context Likelihood of Relatedness (CLR)                | 41        |
| 3.5.1    | What it Does                                           | 41        |
| 3.5.2    | The Data                                               | 41        |
| 3.5.3    | The Strategy                                           | 41        |
| 3.5.4    | Performance on Examples                                | 42        |
| 3.6      | Semidefinite Programming                               | 46        |
| 3.6.1    | What it does                                           | 46        |
| 3.6.2    | The Data                                               | 46        |
| 3.6.3    | The Strategy                                           | 46        |
| 3.6.4    | Performance on Examples                                | 48        |
| <b>4</b> | <b>Step 3: Using Time-Series Data</b>                  | <b>51</b> |
| 4.1      | Introduction                                           | 51        |
| 4.2      | Time-Delay ARACNE                                      | 51        |
| 4.2.1    | What It Does                                           | 51        |
| 4.2.2    | The Data                                               | 52        |
| 4.2.3    | The Strategy                                           | 52        |
| 4.2.4    | Performance on Examples                                | 54        |
| 4.3      | Time-Lagged Context Likelihood of Relatedness (tlCLR)  | 56        |
| 4.3.1    | What it Does                                           | 56        |
| 4.3.2    | The Data                                               | 56        |
| 4.3.3    | The Strategy                                           | 56        |
| 4.3.4    | Performance on Examples                                | 58        |
| 4.4      | Inferelator                                            | 62        |
| 4.4.1    | What it Does                                           | 62        |
| 4.4.2    | The Data                                               | 62        |
| 4.4.3    | The Strategy                                           | 62        |
| 4.4.4    | Performance on examples                                | 64        |
| 4.5      | Dynamic Factor Graphs (DFG)                            | 65        |
| 4.5.1    | What it does                                           | 65        |
| 4.5.2    | The Data                                               | 65        |
| 4.5.3    | The Strategy                                           | 65        |
| 4.5.4    | Performance on examples                                | 67        |
| 4.6      | Bayesian Network Inference with Java Objects (BANJO)   | 72        |

|          |                                                            |           |
|----------|------------------------------------------------------------|-----------|
| 4.6.1    | What it does                                               | 72        |
| 4.6.2    | The Data                                                   | 72        |
| 4.6.3    | The Strategy                                               | 73        |
| 4.6.4    | Performance on Examples                                    | 74        |
| <b>5</b> | <b>Step 4: Pipelines</b>                                   | <b>77</b> |
| 5.1      | Consensus Step: Combining Results of Different Approaches  | 77        |
| 5.2      | Creating Pipelines and Ensemble Networks                   | 77        |
| 5.3      | Pipeline 1: Steady State algorithm + Dynamic Factor Graphs | 78        |
| 5.3.1    | What it does                                               | 78        |
| 5.3.2    | The Strategy                                               | 78        |
| 5.3.3    | Performance on examples                                    | 78        |
| 5.4      | Pipeline 2: Inferelator 2.0                                | 82        |
| 5.4.1    | What it does                                               | 82        |
| 5.4.2    | The Strategy                                               | 82        |
| 5.4.3    | Performance on examples                                    | 83        |
| 5.5      | Ensemble 1: Voting                                         | 86        |
| 5.5.1    | What it does                                               | 86        |
| 5.5.2    | The Strategy                                               | 86        |
| 5.5.3    | Performance on examples                                    | 87        |
| 5.6      | Ensemble 2: Simulated Annealing                            | 90        |
| 5.6.1    | What it does                                               | 90        |
| 5.6.2    | The Data                                                   | 90        |
| 5.6.3    | The Strategy                                               | 90        |
| 5.6.4    | Performance on examples                                    | 91        |
| 5.7      | Conclusions                                                | 95        |
|          | References                                                 | 97        |
|          | <b>Index</b>                                               | <b>99</b> |