

Spring 5-31-2015

Unsupervised gene regulatory network inference on microarray data

Nidhi Radia
New Jersey Institute of Technology

Follow this and additional works at: <https://digitalcommons.njit.edu/theses>



Part of the [Bioinformatics Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

Radia, Nidhi, "Unsupervised gene regulatory network inference on microarray data" (2015). *Theses*. 242.
<https://digitalcommons.njit.edu/theses/242>

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at Digital Commons @ NJIT. It has been accepted for inclusion in Theses by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

UNSUPERVISED GENE REGULATORY NETWORK INFERENCE ON MICROARRAY DATA

**by
Nidhi Radia**

Obtaining gene regulatory networks (GRNs) from expression data is a challenging and crucial task. Many computational methods and algorithms have been developed to infer gene networks for gene expression data, which are usually obtained from microarray experiments. A gene network is a method to depict the relation among clusters of genes. To infer gene networks, the unsupervised method is used in this study. The two types of data used are time-series data and steady-state data. The data is analyzed using various tools containing different algorithms and concepts. GRNs from time-series data tools are obtained using the Time-delayed Algorithm for the Reconstruction of Accurate Cellular Networks (TD-ARACNe), the Bayesian Network Inference with Java Objects (BANJO), and causality. For steady-state data tools such as ARACNe, Gene Network Inference with Ensemble of trees (GENIE3), Context Likelihood or Relatedness Network (CLR), and Maximum Relevance Minimum Redundancy (MRNET) are used. The performance of time-series data as well as steady-state data based tool algorithms is compared by calculating their accuracy. The accuracy is calculated by comparing gene interactions between predicted and true networks. From the experimental studies it was found that the TD-ARACNe gives the highest accuracy on time-series gene expression data while for steady-state data, the ARACNe tool gives the highest accuracy. Overall, these analyses suggest that the suitability of the tools depends on the types of gene expression data available.

**UNSUPERVISED GENE REGULATORY NETWORK INFERENCE ON
MICROARRAY DATA**

**by
Nidhi Radia**

**A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics**

Department of Computer Science

May 2015

APPROVAL PAGE

**UNSUPERVISED GENE REGULATORY NETWORK INFERENCE ON
MICROARRAY DATA**

Nidhi Radia

Dr. Jason T.L. Wang, Thesis Advisor Date
Professor of Bioinformatics and Computer Science, NJIT

Dr. Usman Roshan, Committee Member Date
Associate Professor of Computer Science, NJIT

Dr. Dimitri Theodoratos, Committee Member Date
Associate Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Nidhi Radia

Degree: Master of Science

Date: May 2015

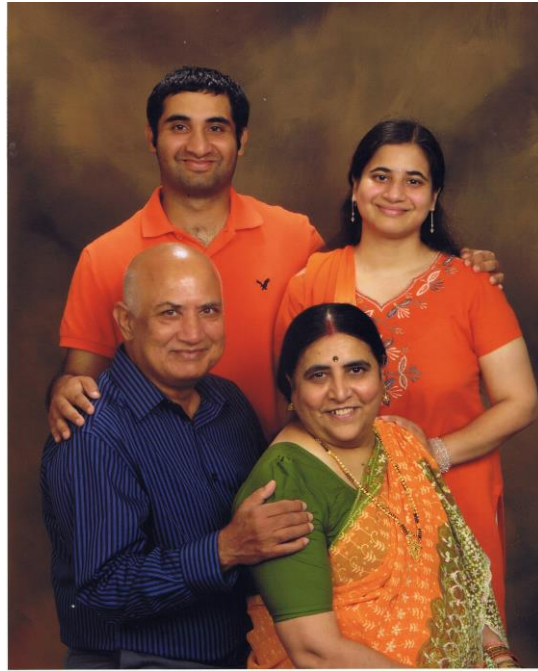
Undergraduate and Graduate Education:

- Master of Science in Bioinformatics,
New Jersey Institute of Technology, Newark, NJ, 2015
- Bachelor of Science in Biotechnology,
Rutgers University, New Brunswick, NJ, 2011

Major: Bioinformatics



*Dedicated to
My Gurus - My Parents – Hansa Radia and Nilesh Radia
My Little One- My Brother – Ankur Radia*



ACKNOWLEDGMENTS

As I am about to finish my MS degree, I realize that my success as a graduate student is due to the unconditional support of my family, advisors, professors, friends, co-workers and the Almighty. Thank you for being part of my journey.

I am highly obliged to Dr. Jason Wang for being my thesis advisor and constantly guiding me throughout the project. His consistent mentoring, valuable time and patience helped me to learn throughout my academic career at NJIT. I will always cherish the liberty that he granted me in the lab facilities by providing me a workspace. I would like to express my deepest gratitude to Dr. Usman Roshan and Dr. Dimitri Theodoratos for being on my Master's thesis committee. Their knowledge and discussions have always helped me to look at my research problem differently.

I am indebted to my employer Bio-reference Laboratory Inc. and my supervisor Ms. Linda Cilindrello without whose support this thesis would have never been possible. I would like to thank my director Mr. Frank Buccini and my co-workers for constant encouragement and keeping my spirits high all the time. I would also like to express my gratitude to Dr. Sun Yi for her discussions on research which inspired me great deal. I am also thankful to Dr. Lata Govada for always being there to guide and help me regarding professional writing such as research papers. I really appreciate Lydia Paradiso's contribution for creating the tutorial on Bioinformatics which helped me greatly while writing my thesis.

During my two years at NJIT, I had the opportunity to meet and interact with dozens of students from the around world; some of them became my extended family and left footprints in my life. I will always cherish the fond memories that I have shared with

Nihir Patel, Akhila Nagula, and Manan Amin whom I lovingly call my bioinformatics gang. Thank you for always being there whenever I needed help and motivating me throughout. I would take this opportunity to thank my DeepCS club and its E-board members for encouraging me to be part of it and allowing me to organize amazing events on campus.

Last, thanks to my beloved family for their sacrifices, eternal love, unconditional support and encouragement to continue my education. Thank you for always being there in my ups and downs, and supporting me morally.

Thank you very much, everyone, for being part of my amazing yet beautiful journey and helping me to achieve my goals.

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION.....	1
1.1 Objective.....	1
1.2 Background Information.....	1
2 DATASET.....	3
2.1 Microarray Data.....	3
2.1.1 Types of Microarray Dataset.....	3
2.2 Time-Series Dataset.....	4
2.3 Steady-State Dataset.....	7
2.4 Dataset Pre-Processing.....	9
2.4.1 R Studio.....	9
3 METHODS.....	11
3.1 Cytoscape.....	11
3.2 Time-Series Tools.....	11
3.2.1 TD-ARACNe.....	11
3.2.2 BANJO.....	13
3.2.3 Causality.....	12
3.3 Steady-State Tools.....	14
3.3.1 ARACNe.....	14
3.3.2 GENIE3.....	16
3.3.3 CLR.....	17

TABLE OF CONTENTS
(Continued)

Chapter	Page
3.3.4 MRNET.....	17
4 RESULTS AND DISCUSSION.....	19
4.1 Comparison of Network Obtained from Time-Series Tools.....	19
4.1.1 TD-ARACNe.....	19
4.1.2 BANJO.....	22
4.2.3 Causality.....	25
4.2 Comparison of Network Obtained from Steady-State Tools.....	29
4.2.1 ARACNe.....	29
4.2.2 GENIE3.....	39
4.2.3 CLR.....	42
4.2.4 MRNET.....	51
4.3 Comparison of Time-Series Tools Accuracy.....	62
4.3.1 TD-ARACNe.....	62
4.3.2 BANJO.....	62
4.3.2 Causality.....	63
4.4 Comparison of Steady-State Tools Accuracy.....	63
4.4.1 ARACNe.....	63

TABLE OF CONTENTS
(Continued)

Chapter	Page
4.4.2 GENIE3.....	64
4.4.3 CLR.....	64
4.4.4 MRNET.....	65
5 CONCLUSION.....	67
REFERENCES.....	68

LIST OF TABLES

Table	Page
2.1 DREAM4 size 10 Network 1 Time-Series Data.....	5
2.2 DREAM4 size 10 Network 2 Time-Series Data.....	5
2.3 DREAM4 size 10 Network 3 Time-Series Data.....	6
2.4 DREAM4 size 10 Network 4 Time-Series Data.....	6
2.5 DREAM4 size 10 Network 5 Time-Series Data.....	7
2.6 DREAM4 size 10 Network 1 Multifactorial Steady-State Data.....	8
2.7 DREAM4 size 10 Network 2 Multifactorial Steady-State Data	8
2.8 DREAM4 size 10 Network 1 Multifactorial Steady-State Data	8
2.9 DREAM4 size 10 Network 1 Multifactorial Steady-State Data	9
2.10 DREAM4 size 10 network 1 Multifactorial Steady-State Data	9
4.1 Accuracy of TD-ARACNe Tool Obtained from DREAM4 Dataset.....	62
4.2 Accuracy of BANJO Tool Obtained from DREAM4 Dataset.....	62
4.3 Accuracy of Causality Tool Obtained from DREAM4 Dataset.....	63
4.4 Accuracy of ARACNe Tool using Kendell Algorithm Obtained from DREAM4.	63
4.5 Accuracy of ARACNe Tool using Pearson Algorithm Obtained from DREAM4.....	63
4.6 Accuracy of ARACNe Tool using Spearman Algorithm Obtained from DREAM4.....	64
4.7 Accuracy of GENIE3 Tool Obtained from DREAM4 Dataset.....	64

LIST OF TABLES

Table	Page
4.8 Accuracy of CLR Tool using Kendell Algorithm Obtained from DREAM4.....	64
4.9 Accuracy of CLR Tool using Pearson Algorithm Obtained from DREAM4.....	64
4.10 Accuracy of CLR Tool using Spearman Algorithm Obtained from DREAM4....	65
4.11 Accuracy of MRNET Tool using Kendell Algorithm Obtained from DREAM4..	65
4.12 Accuracy of MRNET Tool using Pearson Algorithm Obtained from DREAM4..	65
4.13 Accuracy of MRNET Tool using Spearman algorithm Obtained from DREAM4	65

LIST OF FIGURES

Figure	Page
4.1 DREAM4 size 10 network 1 obtained by TD-ARACNe	19
4.2 DREAM4 size 10 network 2 obtained by TD-ARACNe	20
4.3 DREAM4 size 10 network 3 obtained by TD-ARACNe	20
4.4 DREAM4 size 10 network 4 obtained by TD-ARACNe	21
4.5 DREAM4 size 10 network 5 obtained by TD-ARACNe	21
4.6 DREAM4 size 10 network 1 produced by BANJO tool	22
4.7 DREAM4 size 10 network 2 produced by BANJO tool	23
4.8 DREAM4 size 10 network 3 produced by BANJO tool	23
4.9 DREAM4 size 10 network 4 produced by BANJO tool	24
4.10 DREAM4 size 10 network 5 produced by BANJO tool	25
4.11 Gene inference of DREAM4 size 10 network 1 by causality algorithm.....	25
4.12 Gene inference of DREAM4 size 10 network 2 by causality algorithm.....	26
4.13 Gene inference of DREAM4 size 10 network 3 by causality algorithm.....	26
4.14 Gene inference of DREAM4 size 10 network 4 by causality algorithm.....	27
4.15 Gene inference of DREAM4 size 10 network 5 by causality algorithm.....	28
4.16 DREAM4 size 10 network 1 created by ARACNe using Kendell algorithm.....	29
4.17 DREAM4 size 10 network 1 created by ARACNe using Pearson algorithm.....	30
4.18 DREAM4 size 10 network 1 created by ARACNe using Spearman algorithm....	30
4.19 DREAM4 size 10 network 2 created by ARACNe using Kendell algorithm.....	31

LIST OF FIGURES
(Continued)

Figure	Page
4.20 DREAM4 size 10 network 2 created by ARACNe using Pearson algorithm.....	32
4.21 DREAM4 size 10 network 2 created by ARACNe using Spearman algorithm....	32
4.22 DREAM4 size 10 network 3 created by ARACNe using Kendell algorithm.....	33
4.23 DREAM4 size 10 network 3 created by ARACNe using Pearson algorithm.....	33
4.24 DREAM4 size 10 network 3 created by ARACNe using Spearman algorithm....	34
4.25 DREAM4 size 10 network 4 created by ARACNe using Kendell algorithm.....	35
4.26 DREAM4 size 10 network 4 created by ARACNe using Pearson algorithm.....	35
4.27 DREAM4 size 10 network 4 created by ARACNe using Spearman algorithm....	36
4.28 DREAM4 size 10 network 5 created by ARACNe using Kendell algorithm.....	37
4.29 DREAM4 size 10 network 5 created by ARACNe using Pearson algorithm.....	37
4.30 DREAM4 size 10 network 5 created by ARACNe using Spearman algorithm....	38
4.31 DREAM4 size 10 network 1 built by GENIE3.....	39
4.32 DREAM4 size 10 network 2 built by GENIE3.....	39
4.33 DREAM4 size 10 network 3 built by GENIE3.....	40
4.34 DREAM4 size 10 network 4 built by GENIE3.....	41
4.35 DREAM4 size 10 network 5 built by GENIE3.....	42
4.36 DREAM4 size 10 network 1 generated by CLR using Kendell algorithm.....	42
4.37 DREAM4 size 10 network 1 generated by CLR using Pearson algorithm.....	43
4.38 DREAM4 size 10 network 1 generated by CLR using Spearman algorithm.....	43

LIST OF FIGURES
(Continued)

Figure	Page
4.39 DREAM4 size 10 network 2 generated by CLR using Kendell algorithm.....	44
4.40 DREAM4 size 10 network 2 generated by CLR using Pearson algorithm.....	44
4.41 DREAM4 size 10 network 2 using generated by CLR using Spearman algorithm	45
4.42 DREAM4 size 10 network 3 generated by CLR using Kendell algorithm.....	46
4.43 DREAM4 size 10 network 3 generated by CLR using Pearson algorithm.....	46
4.44 DREAM4 size 10 network 3 generated by CLR using Spearman algorithm.....	47
4.45 DREAM4 size 10 network 4 generated by CLR using Kendell algorithm.....	47
4.46 DREAM4 size 10 network 4 generated by CLR using Pearson algorithm.....	48
4.47 DREAM4 size 10 network 4 generated by CLR using Spearman algorithm.....	49
4.48 DREAM4 size 10 network 5 generated by CLR using Kendell algorithm.....	49
4.49 DREAM4 size 10 network 5 generated by CLR using Pearson algorithm.....	50
4.50 DREAM4 size 10 network 5 generated by CLR using Spearman algorithm.....	50
4.51 Gene network of DREAM4 size 10 network 1 using Kendell method by MRNET.....	51
4.52 Gene network of DREAM4 size 10 network 1 using Pearson method by MRNET.....	52
4.53 Gene network of DREAM4 size 10 network 1 using Spearman method by MRNET.....	52
4.54 Gene network of DREAM4 size 10 network 2 using Kendell method by MRNET.....	53

LIST OF FIGURES
(Continued)

Figure	Page
4.55 Gene network of DREAM4 size 10 network 2 using Pearson method by MRNET.....	54
4.56 Gene network of DREAM4 size 10 network 2 using Spearman method by MRNET.....	54
4.57 Gene network of DREAM4 size 10 network 3 using Kendell method by MRNET.....	55
4.58 Gene network of DREAM4 size 10 network 3 using Pearson method by MRNET.....	56
4.59 Gene network of DREAM4 size 10 network 3 using Spearman method by MRNET.....	57
4.60 Gene network of DREAM4 size 10 network 4 using Kendell method by MRNET.....	58
4.61 Gene network of DREAM4 size 10 network 4 using Pearson method by MRNET.....	58
4.62 Gene network of DREAM4 size 10 network 4 using Spearman method by MRNET.....	59
4.63 Gene network of DREAM4 size 10 network 5 using Kendell method by MRNET.....	60
4.64 Gene network of DREAM4 size 10 network 5 using Pearson method by MRNET.....	60
4.65 Gene network of DREAM4 size 10 network 5 using Spearman method by MRNET.....	61

LIST OF ABBREVIATIONS

ARACNe	Algorithm for the Reconstruction of Accurate Cellular Networks
BANJO	Bayesian Network Inference with Java Objects
TDARACNe	TimeDelay-ARACNE
CLR	Context Likelihood or Relatedness Network
MRNET	Maximum Relevance Minimum Redundancy
GRNs	Gene Regulatory Networks
GENIE3	Gene Network Inference with Ensemble of trees
DREAM	Dialogue for Reverse Engineering Assessments and Methods

CHAPTER 1

INTRODUCTION

1.1 Objective

The objective of this thesis is to find the accuracy of bioinformatics tools used for obtaining gene regulatory networks (GRNs). This thesis primarily focused on two types of microarray data, namely time-series data and steady-state data.

For time-series data, tools such as TD-ARACNe, BANJO, and causality are used to infer networks. Then results are analyzed using golden standard data to find the accuracy of the tools, and the ranking of tools is decided later on.

For steady-state data, ARACNe, GENIE3, CLR and MRNET are used to produce GRNs. Afterwards, results are analyzed in the same manner as the time-series results.

1.2 Background Information

Gene networking is a method to depict clusters of genes and the relations among them. Many computational methods and algorithms have been developed to infer gene networks for gene expression data which is usually obtained from microarray experiments. Obtaining GRN models is quicker and cheaper than lab experiments GRNs is a model consisting a network which is composed of nodes and edges. The nodes represent genes, proteins or metabolites, and the edges comprise of molecular interactions such as between proteins and DNA, between two genes or between two proteins. Another important role of GRNs in system biology is to help to understand life processes such as cell differentiation, metabolism, the cell cycle, and signal transduction. This role further

helps to give better ideas about mechanisms of diseases that occur when these cellular processes are dysregulated. Hence, these mechanisms give more insight on gene-pathways, phenotype expression, and genes responsible for cancer. For example, understanding the interaction between genes can help to identify molecular targets for specific drugs, or drugs for specific targets. This knowledge combined with more information on the network behavior and biological models helps to find disease-specific cures and design drugs for personalized medicines.

In this thesis, GRNs are deduced between two genes using microarray data. Two types of microarray data are used namely, time-course experiments, and perturbation experiments. The method used to infer GRNs is unsupervised.

CHAPTER 2

DATASET

In this Chapter, the datasets and their types will be discussed. Later in this chapter, focus will be on the description of data used for the thesis and how they were processed using R studio.

2.1 Microarray Data

Microarray is a method where thousands of genes are expressed using probes that are complementary to the sequences of either DNA or RNA. It obtains raw data in the form of a matrix. Hence, data are required to be pre-processed data to analyze similarities and differences of the genes expressed in the experiment. Hence, the microarray data is also known as gene expression profile. Gene expression means the amount of a gene expressed in given conditions. Usually the gene expression matrix is composed of rows and columns where genes are present in the row while the columns represent conditions such as an array, gene-chip, or experiments. Experiments usually correspond to time-points for time-series dataset, and to conditions or treatments for steady-state dataset.

2.1.1 Types of Microarray Datasets

- 1) Knockdown: Knockdown data is a type of expression data where one or more genes are reduced. Genes can be reduced either by genetic modification or by treating genes with reagents.
- 2) Knockout: In this type of data, one or more genes have been knocked out, meaning genes are being inactivated. In this type of data, inactivation of genes can be done either by replacing or disrupting DNA piece artificially. Hence, the loss of genes changes characteristics such as phenotypes.

- 3) Wildtype: Wildtype data are expression values from the most common phenotype of a given organism.
- 4) Multifactorial: In this type of data more than one perturbation is applied to a set of genes. It is steady-state data meaning expression value are not changing or are changing very slowly. Perturbation can also be applied to either all genes or some genes in the experiments.
- 5) Time-series: In time series data, genetic or chemical perturbation is applied to a set of genes, and readings are noted while expression is changing in the experiment.

2.2 Time-Series Dataset

Dialogue for **R**everse **E**ngineering **A**ssessments and **M**ethods (DREAM) challenge 4 data were used for time-series tool analysis. DREAM4 data was created for experiments and theories in the area of cellular network inference and quantitative model building in systems biology. This data have 10 genes with 21 different time points. In the data matrix, the rows represent different time points while the columns contain various genes. It has five different sets of network data. The initial condition in the data is considered as a steady-state measurement of the wild-type. A perturbation is applied at time $(t) = 0$ minutes (mins). From $t = 0$ mins to $t = 500$ mins, meaning the first half of the time series matrix, the results show the response of the gene expression to the perturbation that was applied. After recording the responses, the perturbation is removed at $t = 500$ mins for the second half of the time series till the last time point $t = 1000$ mins, the results show how the gene expression levels go back from the perturbed to the wild-type state. In short, the wild-type network is restored again in the expression profile.

In the following Tables from 2.1 to 2.5, G represents various genes, and various time- points are measured in minutes.

Table 2.1 DREAM4 size 10 Network 1 Time-Series Data

Time	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0	0.6665	0.1272	0.3551	0.7746	0.1004	0.2755	0.6068	0.7431	0.6656	0.6951
50	0.3258	0.1218	0.3464	0.7229	0.1925	0.3108	0.6097	0.7568	0.5554	0.7327
100	0.1775	0.0444	0.5713	0.5868	0.2333	0.3570	0.4647	0.6657	0.7211	0.6717
150	0.1839	0.0615	0.4850	0.6338	0.4045	0.3376	0.4200	0.7324	0.6402	0.6970
200	0.0931	0.1398	0.3436	0.5354	0.5583	0.2994	0.4145	0.6771	0.5479	0.7513
250	0.0653	0.0886	0.3124	0.5256	0.6429	0.3677	0.5789	0.8026	0.5983	0.6089
300	0.1502	0.0948	0.4126	0.4600	0.5350	0.2993	0.5994	0.6182	0.7289	0.6660
350	0.0913	0.0905	0.3765	0.4214	0.5642	0.3430	0.6775	0.5235	0.7737	0.7444
400	0.1359	0.1182	0.5189	0.5030	0.6719	0.2333	0.5143	0.6088	0.6844	0.8365
450	0.1667	0.0766	0.3769	0.4831	0.6740	0.3816	0.4996	0.5959	0.7487	0.7173
500	0.1170	0.0755	0.4322	0.5403	0.7355	0.3345	0.5091	0.7153	0.6726	0.6326
550	0.4709	0.0937	0.3453	0.6184	0.4853	0.3585	0.4479	0.7052	0.7919	0.5868
600	0.5545	0.1443	0.4763	0.6234	0.3258	0.3239	0.4197	0.7299	0.5792	0.6030
650	0.5718	0.1580	0.3528	0.6429	0.2634	0.2929	0.4525	0.7211	0.6633	0.6928
700	0.6437	0.0906	0.3818	0.7448	0.1907	0.3050	0.6357	0.6836	0.8094	0.7287
750	0.6989	0.0959	0.4026	0.6478	0.1278	0.2935	0.5680	0.7436	0.6820	0.8214
800	0.6002	0.1251	0.3520	0.7884	0.1340	0.3429	0.5652	0.6701	0.7549	0.6814
850	0.7517	0.0985	0.4542	0.7087	0.1392	0.3147	0.3950	0.6563	0.6583	0.6836
900	0.7058	0.0824	0.4255	0.7010	0.0960	0.3309	0.4628	0.6610	0.6866	0.7454
950	0.6863	0.1131	0.3168	0.6602	0.1390	0.3387	0.5904	0.6152	0.7002	0.6451
1000	0.6786	0.0781	0.3340	0.6695	0.1182	0.2846	0.5939	0.6824	0.7702	0.6068

Table 2.2 DREAM4 size 10 Network 2 Time-Series Data

Time	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0	0.6629	0.1479	0.6120	0.7147	0.1672	0.6486	0.0417	0.7280	0.6067	0.5978
50	0.6762	0.2663	0.5867	0.6211	0.1768	0.2115	0.0958	0.7279	0.6161	0.3535
100	0.6499	0.2271	0.5552	0.7286	0.2841	0.1475	0.1594	0.6161	0.6102	0.2467
150	0.7366	0.3119	0.6120	0.6364	0.2322	0.1166	0.1082	0.6998	0.6555	0.2297
200	0.6697	0.3594	0.5522	0.6019	0.1690	0.0843	0.0391	0.6284	0.6063	0.2421
250	0.7655	0.3480	0.5896	0.8418	0.2197	0.0368	0.0302	0.5903	0.5750	0.2617
300	0.6713	0.3012	0.5642	0.6612	0.1955	0.0546	0.0192	0.4982	0.6052	0.2322
350	0.6531	0.3964	0.5156	0.6545	0.2205	0.0567	0.0101	0.5552	0.8884	0.2834
400	0.7095	0.4858	0.7352	0.6959	0.2654	0.0984	0.0140	0.5764	0.5945	0.3155
450	0.6442	0.3658	0.6183	0.7552	0.1891	0.0743	0.0121	0.5709	0.5913	0.2744
500	0.7321	0.3768	0.5533	0.7484	0.1736	0.0686	0.0209	0.4771	0.6506	0.3445
550	0.7683	0.2599	0.5805	0.6934	0.0679	0.4207	0.0564	0.5144	0.6625	0.6782
600	0.7135	0.0946	0.6226	0.7316	0.0502	0.5567	0.1188	0.6730	0.6116	0.6946
650	0.5405	0.0657	0.7016	0.6596	0.0637	0.4946	0.1145	0.6814	0.7677	0.6962
700	0.7235	0.0698	0.6252	0.7474	0.1991	0.5933	0.0661	0.6835	0.6244	0.6351
750	0.7445	0.1791	0.6708	0.7099	0.2486	0.5812	0.0688	0.7071	0.8015	0.7254
800	0.6565	0.2643	0.5605	0.8040	0.1119	0.6017	0.1092	0.6559	0.5745	0.7505
850	0.6644	0.1535	0.6601	0.7343	0.0559	0.7040	0.1106	0.6533	0.5741	0.6199
900	0.5947	0.0480	0.7683	0.7035	0.0614	0.7472	0.0781	0.6549	0.6811	0.6628
950	0.7603	0.0464	0.6377	0.6772	0.1621	0.5347	0.0546	0.7131	0.5368	0.7113
1000	0.5853	0.1089	0.5751	0.7424	0.1135	0.6229	0.0561	0.7519	0.5974	0.6970

Table 2.3 DREAM4 size 10 Network 3 Time-Series Data

Time	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0	0.402	0.182	0.619	0.522	0.491	0.291	0.151	0.69	0.65	0.684
50	0.436	0.436	0.512	0.493	0.5	0.275	0.164	0.648	0.796	0.321
100	0.583	0.523	0.482	0.453	0.498	0.145	0.102	0.597	0.676	0.13
150	0.647	0.645	0.511	0.496	0.408	0.192	0.049	0.394	0.665	0.234
200	0.636	0.581	0.473	0.43	0.63	0.183	0.073	0.255	0.658	0.192
250	0.602	0.52	0.446	0.376	0.557	0.185	0.12	0.246	0.664	0.207
300	0.664	0.547	0.472	0.344	0.619	0.133	0.106	0.258	0.864	0.154
350	0.608	0.745	0.463	0.385	0.532	0.168	0.119	0.333	0.872	0.161
400	0.695	0.639	0.438	0.361	0.549	0.181	0.143	0.31	0.638	0.139
450	0.555	0.627	0.432	0.259	0.534	0.14	0.084	0.317	0.583	0.165
500	0.643	0.64	0.375	0.288	0.66	0.161	0.068	0.259	0.682	0.265
550	0.721	0.394	0.297	0.356	0.656	0.144	0.067	0.177	0.749	0.492
600	0.681	0.307	0.454	0.334	0.513	0.173	0.149	0.384	0.743	0.701
650	0.506	0.32	0.49	0.373	0.607	0.161	0.213	0.413	0.803	0.72
700	0.476	0.35	0.449	0.523	0.539	0.216	0.229	0.573	0.604	0.68
750	0.33	0.4	0.423	0.513	0.604	0.279	0.252	0.667	0.616	0.695
800	0.28	0.357	0.458	0.471	0.591	0.179	0.186	0.688	0.682	0.839
850	0.45	0.346	0.483	0.549	0.669	0.233	0.157	0.682	0.743	0.649
900	0.572	0.33	0.542	0.682	0.666	0.189	0.194	0.627	0.642	0.788
950	0.388	0.298	0.48	0.52	0.583	0.264	0.185	0.683	0.786	0.72
1000	0.455	0.337	0.512	0.629	0.456	0.28	0.302	0.548	0.69	0.74

Table 2.4 DREAM4 size 10 Network 4 Time-Series Data

Time	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0	0.557	0.197	0.581	0.17	0.531	0.623	0.39	0.739	0.691	0.675
50	0.501	0.204	0.757	0.222	0.656	0.678	0.553	0.619	0.763	0.912
100	0.313	0.175	0.653	0.188	0.687	0.705	0.692	0.609	0.66	0.554
150	0.238	0.184	0.685	0.191	0.595	0.639	0.599	0.716	0.68	0.587
200	0.222	0.356	0.651	0.18	0.736	0.609	0.639	0.669	0.694	0.73
250	0.191	0.381	0.626	0.19	0.615	0.732	0.628	0.788	0.733	0.672
300	0.188	0.283	0.404	0.188	0.757	0.651	0.794	0.696	0.692	0.757
350	0.224	0.427	0.303	0.183	0.65	0.625	0.628	0.718	0.584	0.657
400	0.223	0.398	0.304	0.206	0.597	0.637	0.671	0.73	0.479	0.655
450	0.246	0.333	0.31	0.275	0.686	0.637	0.726	0.646	0.539	0.703
500	0.162	0.385	0.343	0.255	0.748	0.692	0.7	0.657	0.668	0.719
550	0.314	0.443	0.334	0.178	0.76	0.592	0.437	0.882	0.545	0.721
600	0.262	0.354	0.363	0.303	0.858	0.912	0.319	0.747	0.524	0.781
650	0.368	0.343	0.345	0.205	0.664	0.727	0.252	0.716	0.627	0.699
700	0.499	0.24	0.335	0.235	0.683	0.707	0.332	0.62	0.637	0.726
750	0.474	0.19	0.404	0.265	0.641	0.788	0.292	0.697	0.748	0.66
800	0.429	0.221	0.482	0.277	0.74	0.726	0.405	0.69	0.635	0.767
850	0.393	0.249	0.649	0.154	0.785	0.759	0.311	0.662	0.674	0.631
900	0.402	0.165	0.639	0.148	0.695	0.804	0.323	0.748	0.706	0.771
950	0.471	0.17	0.668	0.18	0.7	0.723	0.358	0.696	0.738	0.566
1000	0.47	0.181	0.663	0.137	0.581	0.673	0.259	0.685	0.681	0.681

Table 2.5 DREAM4 size 10 Network 5 Time-Series Data

Time	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0	0.709	0.522	0.631	0.139	0.04	0.073	0.02	0.462	0.21	0.655
50	0.228	0.848	0.562	0.134	0.195	0.043	0.02	0.592	0.198	0.217
100	0.08	0.793	0.479	0.14	0.173	0.168	0.014	0.396	0.18	0.122
150	0.008	0.556	0.506	0.141	0.225	0.294	0.016	0.355	0.107	0.078
200	0.005	0.637	0.429	0.144	0.204	0.438	0.041	0.261	0.192	0.065
250	0.003	0.704	0.388	0.217	0.193	0.515	0.04	0.172	0.252	0.038
300	8E-04	0.653	0.357	0.162	0.195	0.573	0.107	0.119	0.389	0.073
350	0	0.668	0.398	0.233	0.213	0.627	0.12	0.185	0.571	0.093
400	0.002	0.622	0.329	0.14	0.23	0.649	0.121	0.136	0.76	0.07
450	0.002	0.548	0.452	0.222	0.223	0.677	0.145	0.14	0.672	0.05
500	0.005	0.663	0.339	0.199	0.223	0.711	0.095	0.105	0.688	0.092
550	0.171	0.517	0.129	0.239	0.093	0.538	0.171	0.165	0.589	0.661
600	0.272	0.493	0.206	0.512	0.102	0.46	0.102	0.24	0.555	0.627
650	0.277	0.388	0.33	0.428	0.067	0.253	0.083	0.207	0.546	0.779
700	0.368	0.53	0.34	0.171	0.073	0.115	0.055	0.398	0.668	0.801
750	0.338	0.648	0.423	0.2	0.05	0.1	0.038	0.446	0.422	0.804
800	0.543	0.74	0.464	0.166	0.049	0.072	0.042	0.463	0.331	0.693
850	0.683	0.647	0.431	0.168	0.084	0.083	0.014	0.441	0.224	0.697
900	0.694	0.647	0.514	0.098	0.023	0.062	0.062	0.484	0.255	0.662
950	0.744	0.627	0.531	0.158	0.031	0.072	0.02	0.515	0.193	0.655
1000	0.774	0.877	0.426	0.135	0.044	0.063	0.014	0.618	0.281	0.906

2.3 Steady-State Dataset

DREAM4 challenge multifactorial data were used for steady-state tool analyses. The data have 10 genes with 10 different perturbations. In the data matrix, row represents different conditions while column contains genes (G). Different conditions in these data were noted at steady point. The multifactorial data have five different sets of network data. Networks were obtained by applying multifactorial perturbations to the original networks. The Multifactor in the experiments can be different chemicals in different amounts, drugs or other factors. Hence, the conditions affect all genes simultaneously in the experiment.

In the following Tables from 2.6 to 2.10, G represents various genes.

Table 2.6 DREAM4 size 10 Network 1 Multifactorial Steady-State Data

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0.75	0.0145	0.2768	0.5925	0.1137	0.4200	0.6705	0.7873	0.5851	0.6379
0.2	0.0667	0.4668	0.7665	0.5201	0.2496	0.3280	0.7171	0.7696	0.6548
0.77	0.1396	0.1299	0.6305	0.0016	0.2795	0.5784	0.6209	0.5910	0.6678
0.42	0.1072	0.3394	0.6284	0.2198	0.3295	0.6402	0.6728	0.4712	0.8043
0.65	0.0237	0.3495	0.5408	0.0477	0.1340	0.5039	0.4542	0.6125	0.6829
0.22	0.1880	0.3619	0.5945	0.4412	0.3384	0.4117	0.5655	0.6212	0.7914
0.68	0.1318	0.1523	0.7434	0.1429	0.3288	0.6494	0.7462	0.8482	0.6229
0.51	0.1414	0.4101	0.6492	0.0629	0.3404	0.4030	0.6686	0.7033	0.6574
0.68	0.2225	0.4616	0.6675	0.0320	0.6964	0.4223	0.0247	0.7339	0.7638
0.41	0.2649	0.5388	0.7322	0.2661	0.6220	0.1792	0.1225	0.6937	0.8356

Table 2.7 DREAM4 size 10 Network 2 Multifactorial Steady-State Data

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0.6593	0.4455	0.7103	0.5501	0.3063	0.4587	0.3318	0.3076	0.5751	0.783338
0.5714	0.3575	0.6382	0.6608	0.2413	0.4871	0.1773	0.4188	0.7982	0.573654
0.3712	0.2762	0.5225	0.5969	0.2493	0.4129	0.2060	0.5136	0.5473	0.381006
0.6701	0.5445	0.6009	0.7643	0.3548	0.5790	0.4001	0.2215	0.4718	0.708948
0.5559	0.5499	0.7060	0.2935	0.4472	0.3855	0.3462	0.2327	0.5860	0.742855
0.7721	0.4343	0.7696	0.4531	0.3480	0.4478	0.2891	0.2843	0.2297	0.524547
0.6573	0.0975	0.6275	0.6842	0.1837	0.6873	0.0218	0.6287	0.5885	0.585772
0.7384	0.1209	0.5960	0.6310	0.0526	0.5887	0.2022	0.7028	0.8131	0.741225
0.557	0.475	0.58	0.704	0.307	0.435	0.317	0.441	0.272	0.240268
0.518	0.091	0.702	0.657	0.13	0.791	0.012	0.45	0.584	0.589867

Table 2.8 DREAM4 size 10 Network 3 Multifactorial Steady-State Data

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0.19	0.449	0.596	0.156	0.41	0.082	0.45	0.794	0.658	0.826
0.39	0.421	0.505	0.655	0.494	0.16	0.093	0.474	0.404	0.772
0.1	0.416	0.513	0.263	0.651	0.34	0.377	0.755	0.704	0.614
0.14	0.249	0.536	0.303	0.474	0.106	0.555	0.689	0.61	0.771
0.23	0.617	0.218	0.483	0.624	0.047	0.328	0.589	0.463	0.627
0.25	0.526	0.08	0.571	0.564	0.017	0.27	0.68	0.436	0.808
0.08	0.32	0.476	0.524	0.21	0.155	0.279	0.848	0.679	0.485
0.45	0.107	0.711	0.352	0.399	0.382	0.031	0.072	0.778	0.5
0.15	0.591	0.296	0.347	0.565	0.044	0.567	0.663	0.689	0.541
0.1	0.379	0.183	0.649	0.699	0.055	0.341	0.627	0.678	0.683

Table 2.9 DREAM4 size 10 Network 4 Multifactorial Steady-State Data

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0.08	0.583	0.325	0.186	0.567	0.604	0.553	0.507	0.775	0.697
0.4	0.233	0.261	0.289	0.545	0.52	0.329	0.242	0.296	0.649
0.12	0.378	0.576	0.137	0.376	0.638	0.655	0.614	0.804	0.68
0.17	0.306	0.598	0.223	0.571	0.602	0.562	0.622	0.884	0.624
0.78	0.032	0.602	0.179	0.632	0.736	0.336	0.737	0.641	0.51
0.41	0.069	0.392	0.224	0.585	0.549	0.361	0.598	0.421	0.7
0.05	0.626	0.464	0.027	0.85	0.693	0.032	0.512	0.54	0.587
0.18	0.408	0.66	0.121	0.691	0.676	0.493	0.696	0.38	0.477
0.36	0.081	0.694	0.182	0.529	0.592	0.477	0.867	0.65	0.715
0.54	0.039	0.699	0.124	0.638	0.738	0.377	0.567	0.602	0.432

Table 2.10 DREAM4 size 10 Network 5 Multifactorial Steady-State Data

G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
0.2	0.582	0.432	0.169	0.505	0.847	0.104	0.199	0.899	0.656
0.36	0.65	0.522	0.127	0.146	0.208	0.058	0.36	0.446	0.63
0.17	0.237	0.479	0.165	0.367	0.671	0.247	0.092	0.674	0.465
0.24	0.462	0.184	0.349	0.621	0.582	0.235	0.089	0.524	0.756
0.86	0.568	0.466	0.208	0.255	0.802	0.119	0.089	0.196	0.616
0.27	0.506	0.424	0.177	0.458	0.624	0.403	0.087	0.302	0.556
0.51	0.58	0.476	0.062	0.157	0.382	0.322	0.149	0.317	0.659
0.25	0.445	0.263	0.31	0.658	0.608	0.435	0.046	0.721	0.581
0.21	0.654	0.427	0.15	0.015	0.002	0.01	0.748	0.255	0.728
0.15	0.59	0.487	0.021	0.473	0.714	0.117	0.089	0.808	0.655

2.4 Dataset Pre-Processing

Data pre-processing was done using R studio. Most of the tools used for this thesis are R packages which were run and analyzed in R studio.

2.4.1 R studio

R studio was downloaded on a computer. It is an integrated development environment for programming language R, to compute statistical functions and graphics. In this thesis, it

was used to format and analyze data. Many R packages with different algorithms were downloaded for analyzing gene expression data and obtaining output for GRNs.

CHAPTER 3

METHODS

In Chapter 3, methods used for the thesis are discussed. Methods, described in detail, include tools used to infer GRNs.

3.1 Cytoscape

Cytoscape is a bioinformatics tool that helps to visualize molecular interaction network dealing with gene expression data sets. It is an open source software available platform for analyzing biological pathways by implementing these networks with annotations, gene expression profiles, and other data. It is specially designed for biological research, but now it provides a platform to study and illustrate complex gene networks as well. It is a Java-based application which can be easily downloaded on Windows. In this thesis, Cytoscape is used to obtain GRNs as mentioned in sources of all Figures.

3.2 Time-Series Tools

3.2.1 TD-ARACNe

TD-ARACNe is **T**ime **D**elayed**A**lgorithm for the **R**econstruction of **A**ccurate **C**ellular **N**etworks. It is based on dependencies between two genes at different time-points, which consider a measure of these dependencies using mutual information. The main idea of the algorithm is to identify time-delayed dependencies between the expression profiles by undertaking probabilistic model of a stationary Markov Random Field. It infers gene-gene interactions GRNs at certain time expression of gene to previous time point of gene

expression level. Algorithm has three steps in which initially, it detects time-points and expressions of all genes present in the data, then network is constructed, and finally network pruning. The main functions of TD-ARACNe are as follow:

- 1) Z: is the input data in the form of eSet. eSet is the ExpressionSet object. Data were converted into eSet for experimental running.
- 2) N: is number of bins used in normalizations. For this experiment number 100 was used.
- 3) delta: is the maximum time delay allowed for infer connections. In this experiment, value 3 was used for delta to obtain GRN.
- 4) likelihood: It is the fold change used as threshold to state the initial change expression (IcE). For this experiment, value 1 was used for likelihood.
- 5) norm: normalization; Row normalization==1, rank normalization==2. Rank normalization was used in this experiment.
- 6) logarithm: if z is log put logarithm == 0. In this experiment, logarithm 1 was used.
- 7) thresh: it is the Influence threshold. 0 is used for no threshold. Default value 0 was used to run data in the experiment.
- 8) ksd: ksd is the standard deviation multiplier. Value 1 was used for the experiment.
- 9) tolerance: tolerance is the DPI tolerance; 0 for no tolerance;1 for no DPI; 0.15 is the default ARACNE tolerance as it is for TDARACNE. Default value was used for the experiment.
- 10) plot : plot must be TRUE to obtain the graph in R studio.
- 11) dot: returns .dot file in working directory if it's TRUE.
- 12) Adj= TRUE, returns adjacency matrix in RSTUDIO.

Based on above functions commands TD-ARACNe algorithm is run on R studio.

Dot file was obtained as output which was imported into cytoscape to infer networks.

3.2.2 BANJO

BANJO stands for **BA**yesian **N**etwork inference with **J**ava **O**bjects. It is based on statistical concept of Bayesian Network. It is a graphical representation of a joint probability distribution, depicting dependence and conditional independence relationships. BANJO was designed from the ground up to provide efficient structure inference to analyze large research-oriented datasets. Since it is implemented in Java, the framework is easy to maintain and extend. Banjo algorithm does greedy search where it takes input such as gene expression data and gives output in Banjo dot format. It selects random network as it starting point called “Current network”. It is handled by a component called “Proposer” which also proposes new network(s). Based on Bayesian Dirichlet (BD) score, all the neighbors of current Bayesian Network are evaluated which is handled by “Cycle checker.” Another component called “Decider” which decides whether to accept proposed network(s) or not based on best score computed. Lastly, computing the score of the proposed network is handled by “Evaluator” computing. This process continues until none of the neighbors have a BD score higher than the current network. BD score is calculated based on conditional probability distribution that quantifies the effect of parents on nodes. Since BANJO considers Markov Model resulting in conditional independence giving it parents node which is conditionally independent of its non-decedents. Inferred network(s) is built from influence scores. Influence score is defined as metric representing the degree. It helps to predict the sign of networks edges. It gives output file in the directory with network score, influence score and other variables. These scores can be used to infer network in cytoscape to obtain graphical GRNs.

3.2.3 Causality

Causality is defined as the relationship between first event and a second event, where the first event is responsible for the second event. In the gene expression profile, gene A interacts with gene B. The interaction between genes can be either up-regulation or down-regulation. In other words, it also inhibits or promotes transcription in the cell when gene A and gene B expressed. This tool is based on algorithm regarding causal effects. This R package contains implementations of the algorithms PC, FCI, RFCI, GES and GIES, as well as of the IDA method and the generalized Pearl's backdoor criterion. The original PC is the order-dependent, in which the output depends on the order of the given variables. The predefined function in the command used is Gaussian which is a conditional independence test and creates the corresponding sufficient statistic, consisting of the correlation matrix of the data and the sample size. Each conditional independence test can be performed at a certain significance level alpha. Then, the pc function is used to produce an estimate of the underlying causal structure. At the end, results were graphed with directed edges called directed acyclic graph. Similar to TD-ARACNe, commands for this R package is run on R studio and then output was imported into cytoscape to obtain gene network diagrams.

3.3 Steady-State Tools

3.3.1 ARACNe

ARACNe is **A**lgorithm for the **R**econstruction of **A**ccurate **C**ellular **N**etworks. It is an algorithm applied to identify transcriptional interactions between gene products using

microarray gene expression profile data. It requires data samples with data statistically independent from each other. It performs two kinds of calculation as explained below:

- 1) Mutual Information (MI): It compares gene expression profiles using a pairwise correlation coefficient and does not require a monotonic relationship. MI for a pair of random variables, x and y is defined as $MI_{xy} = H_x + H_y - H_{xy}$, where H is the entropy of an arbitrary variable m . It is defined as follow equation

$$H_m = -\sum_{k=1}^n p(x_k) \log(p(x_k)) \quad (3.1)$$

$MI = 0$ if the two variables are statistically independent while $MI > 0$ means that the two variables are non-randomly associated to each other. A computationally efficient Gaussian kernel estimator is used for MI calculation. Threshold I_0 is used to choose interactions and computed for a specific p-value, p_0 .

- 2) Data Processing Inequality (DPI): it calculates MI values of each marker and then removes weakest interaction between markers. If genes g_1 and g_3 interact only through a third gene g_2 , then the following

$$I(g_1, g_3) \leq \min [I(g_1, g_2); I(g_2, g_3)] \quad (3.2)$$

The DPI is applied to triplets of small subsets for whom all three edges can hold and manage the mutual information thresholding. Hence, DPI removes all the indirect interactions.

ARACNE's complexity is $O(N^3 + N^2M^2)$, where M is the number of samples and N is the number of genes. In the equation, N^3 relates to the DPI analysis while N^2M^2 is the mutual information estimation.

Minet R package was used to run ARACNe algorithm with three types of statistical methods are used namely Kendell, Spearman and Pearson. After obtaining the output consists of gene interaction with their weight scores, cytoscape is used for GRNs figures.

3.3.2 GENIE3

It stands for **GE**ne Network Inference with **EN**semble of trees. It is well performed tool for steady-state data. Algorithm contains three steps:

- 1) Initially, it creates an ensemble of regression trees for each gene in the network. It predicts the behavior of each gene in the dataset. In this step, it breaks huge data into small subsets. Based on the threshold value of expressed genes it creates group. Random forest is used to split data and construct regression trees. An important score is calculated for each decision point in the tree. It is defined as a following

$$I(N) = \#S \text{Var}(S) - \#S_t \text{Var}(S_t) - \#S_f \text{Var}(S_f) \quad (3.3)$$

In this formula, N is the decision point; S is the subset of experiment below decision point, S_t and S_f subset of experiments on two branches that is true and false, $\#$ is the number experiments and its subset, var is variance of target gene in a subset.

- 2) Possible regulators are ranked from each regression trees. It can be done by ranking influence of every other gene when tree is created. A score for a potential regulator gene is calculated by summing all of the important scores from the nodes where gene was selected for splitting. Afterwards, based on scores ranking it is determined what genes are important for regulating gene.
- 3) Ranks the inferred edges overall. In this step, important scores are generated for each tree in each forest, which gives list of potential regulators for each tree. Based on tree for each target genes are grouped together then important scores for potential regulators are averaged together and ranked for each genes.

The computational complexity of this tool is $O(pTKN \log N)$ where T is the number of trees, N is the learning sample size, p is the each genes and K is the main parameter of the two tree-based methods.

GENIE3 provides commands for R environment hence it is ran on R studio to obtain the output which later imported into cytoscape for gene network diagram.

3.3.3 CLR

CLR stands for **C**ontext **L**ikelihood of **R**elatedness. CLR calculates the mutual information value from data to infer networks. Mutual information is used as a metric of similarity between two gene expression data. The mutual information score suggests how much information is being shared among genes. For instance, value 0 indicates no information is involved between two genes.

The algorithm of this tool is as follows:

- 1) The mutual information score is calculated between each pair of genes in the data.
- 2) Once the mutual information score is calculated, the score is compared to the background distribution. Background distribution is the row of all mutual information values for two genes. In this step, the likelihood of each mutual information value within its network context is produced. Hence, one will get z scores.
- 3) After the 2nd step is performed, it calculates the CLR score. It is again calculated for each pair of genes by using z scores calculated from the background distribution.

CLR has a complexity in $O(N^2)$ after the mutual information matrix is calculated.

A Minet R package was used to run the CLR algorithm which has three types of statistical methods; Kendell, Spearman, and Pearson. After obtaining the output which consists of gene interactions with their weight scores, Cytoscape is used for GRNs figures.

3.3.4 MRNET

MRNET stands for **M**inimum **R**edundancy **N**ETworks. This method is based on the maximum relevance or minimum redundancy (MRMR), which is an effective information theoretic technique for feature selection. The MRMR principle comprises

selecting the least redundant variables with the highest mutual information towards the target. MRNET uses this principle to infer networks from microarray data. It calculates mutual information similar to the ARACNe algorithm.

The MRNET algorithm selects the variable which has the highest mutual information with its target. Then in the next step, the selected variable chooses the variable that maximizes the scores on the basis of a redundancy concept. The weight of each pair of variables will be the maximum score between the one computed of the one target and the one computed of the second target. The MRNET repeats this selection procedure for each target gene in the set of the gene expression profile. The MRNET algorithm returns a matrix which is the weighted adjacency matrix of the network as output for GRNs.

The complexity of the MRNET algorithm is defined as $O(fxn^2)$ as the feature selection step is repeated for each of n genes in the dataset.

A Minet R package was used to run the MRNET algorithm which has three types of statistical methods; namely Kendall, Spearman, and Pearson. After obtaining the output which, consists of gene interactions with their weight scores, Cytoscape is used for GRNs figures

CHAPTER 4

RESULTS AND DISCUSSION

In this Chapter, results of time-series and steady-state tools will be described. The accuracy of bioinformatics tools will be found and compared with respective types of data.

4.1 Comparison of Network Obtained from Time-series Tools

4.1.1 TD-ARACNe

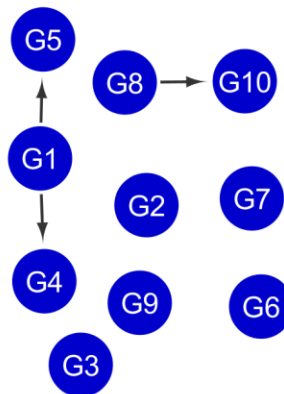


Figure 4.1 DREAM4 size 10 network 1 obtained by TD-ARACNe.
Source: Cytoscape

As seen in Figure 4.1, the GRN of DREAM4 size 10 network 1 was obtained by TD-ARACNe. It shows gene 8 interacts with gene 10. Similarly, gene 1 is interacting with both genes, namely 4 and 5. Genes 2, 3, 4, 6, 7, and 9 are not corresponding with any genes in the network diagram.

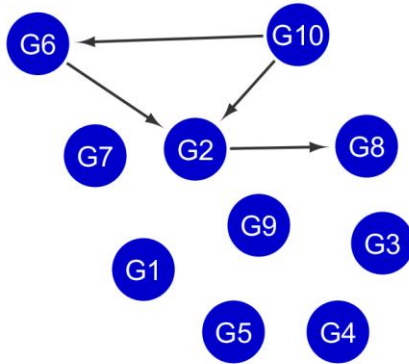


Figure 4.2 DREAM4 size 10 network 2 obtained by TD-ARACNe.
Source: Cytoscape

As seen in Figure 4.2, an inferred network is produced from DREAM4 size 10 network 2 data in which gene 2 is interacting with gene 8. Gene 6 and gene 10 both interface with gene 2. As depicted in the diagram, gene 10 is also communicating with gene 6. Genes such as 1, 3, 4, 5, 7 and 9 are solo ones as clearly seen in the obtained GRN.

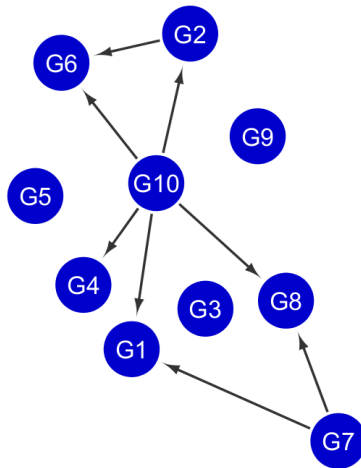


Figure 4.3 DREAM4 size 10 network 3 obtained by TD-ARACNe.
Source: Cytoscape

Gene 10 is interacting with many genes such as genes 1, 2, 4 and 6 as seen in Figure 4.3. Gene 7 is communicating with two gene numbers, namely gene 1 and 8. Also,

the inferred network also depicts that a gene 2 interfaces with gene 6. On the other hand, genes such as 3, 5 and 9 do not interact with any corresponding gene in the diagram.

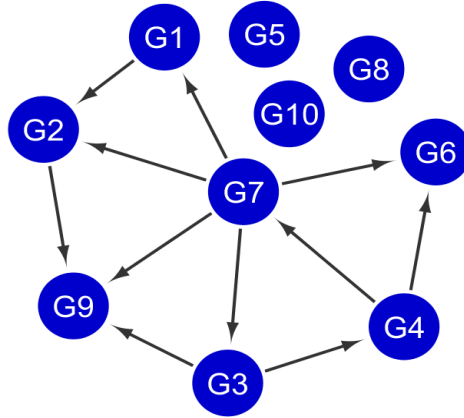


Figure 4.4 DREAM4 size 10 network 4 obtained by TD-ARACNe.
Source: Cytoscape

As seen in Figure 4.4, an inferred gene network was obtained by TD-ARACNe from DREAM4 size 10 network 4 dataset. In the diagram, gene 7 interacts with many genes such as numbers 1, 2, 3, 6, and 9 while genes 5, 8 and 10 do not communicate with any genes. Also, correlation among pairs of genes is between gene 1 with gene 2, gene 2 with gene 9, gene 3 with genes 4 and 9 and lastly gene 4 with genes 6 and 7 as clearly seen in the above GRN.

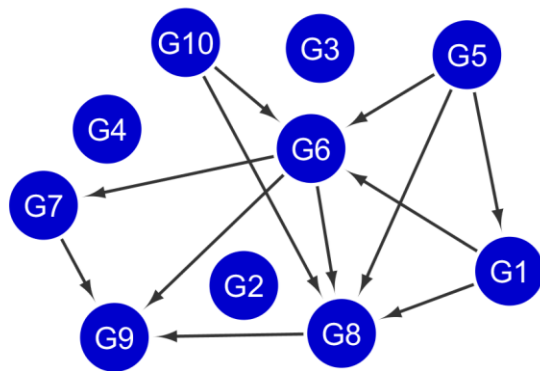


Figure 4.5 DREAM4 size 10 network 5 obtained by TD-ARACNe.
Source: Cytoscape

Figure 4.5 is the inferred network of DREAM4 size 10 network 5 obtained by TD-ARACNe in which genes, namely 2, 3 and 4 do not interact with any genes. Gene numbered 5 is communicating with genes 1, 6 and 8. Also, gene 1 is interacting with two genes 6 and 8. Gene 6 interfaces with three genes numbered 7, 8 and 9. Two genes numbered 7 and 8 only connect with gene 9. Lastly, gene 10 interacts with genes 6 and 8.

4.1.2 BANJO

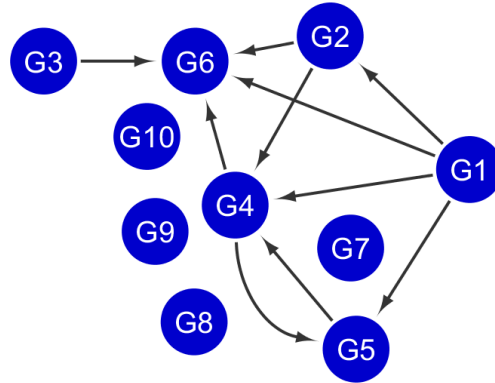


Figure 4.6 DREAM4 size 10 network 1 produced by BANJO tool.

Source: Cytoscape

Figure 4.6 depicts the gene inference network of DREAM4 size 10 network 1 produced by BANJO tool. In this diagram, gene 1 interacts with genes 2, 4, 5, and 6. As seen in Figure 4.6, gene 4 communicates with two genes, namely 5 and 6. Gene 3 only interfaces with gene 6. Similarly, gene 5 connects with gene 4 as they are bidirected. Lastly, gene 2 edges with genes 4 and 6. Gene 7 to gene 10 are solo, meaning not interacting with any corresponded genes.

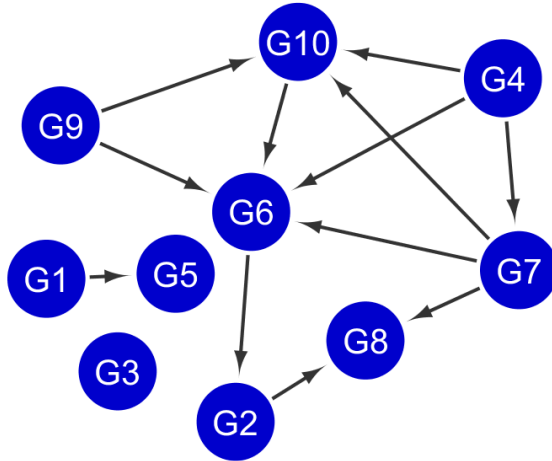


Figure 4.7 DREAM4 size 10 network 2 produced by BANJO tool.
Source: Cytoscape

In this Figure 4.7, Gene 4 is communicating with many genes namely 6, 7 and 10. Only gene 3 is not interacting with any gene. A gene with only a single gene to gene interaction is 2 with 8, 6 with 2, 10 with 6 and 1 with 5. Gene 9 communicates with two genes namely 6 and 10. Lastly, gene 7 interfaces with 6, 8 and 10 as clearly depicted from the GRN obtained by BANJO tool.

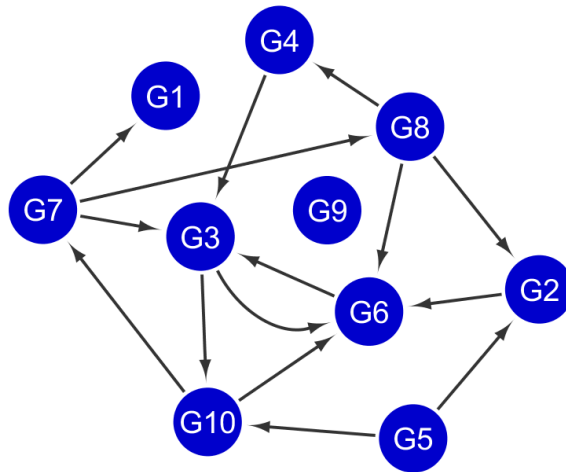


Figure 4.8 DREAM4 size 10 network 3 produced by BANJO tool.
Source: Cytoscape

Figure 4.8 is of DREAM4 size 10 network 3 produced by the BANJO tool. In this diagram, gene 9 is the only gene not interacting with any corresponding genes. Gene 8 is communicating with three genes, namely 2, 4 and 6. In the same manner, gene 7 interacts with gene 1, 3 and 8 as seen in the network. Genes with two interactions are gene 3 with 6 and 10, gene 5 with 2 and 10, and gene 10 with 6 and 7. Lastly solo interfaces among genes pairs are gene 4 with 3, 2 with 6 and 6 with 3.

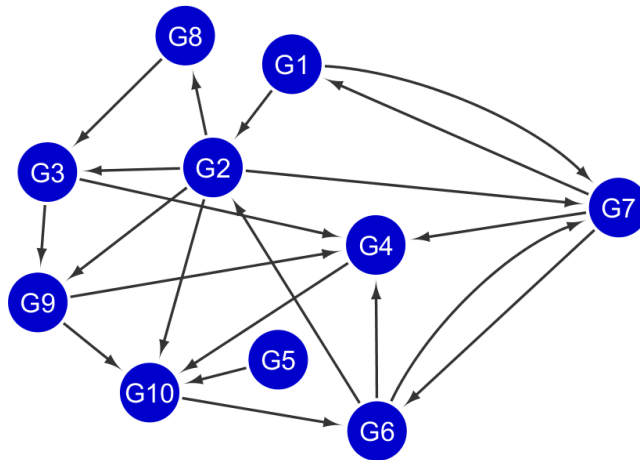


Figure 4.9 DREAM4 size 10 network 4 produced by BANJO tool.
Source: Cytoscape

As seen in Figure 4.9, all genes are interacting to corresponding genes. Gene 7 is bidirected with genes, namely 1 and 6, meaning they are interacting with each other. Hence, birdirected edges are shown in curved arrows as clearly seen in the diagram. Gene 2 is communicating with four genes namely 3, 8, 9 and 10, and gene 1 with gene 2. Genes 6 and 7 also interface with gene 4. Genes such as 4, 5, 8 and 10 interact with only one gene, namely 10, 10, 3 and 6 respectively. As seen in the network, gene 6 communicates with gene 2 and 4. Also, gene 9 connects with gene 10 and 4. Lastly, gene 3 interacts with gene 4 and gene 10.

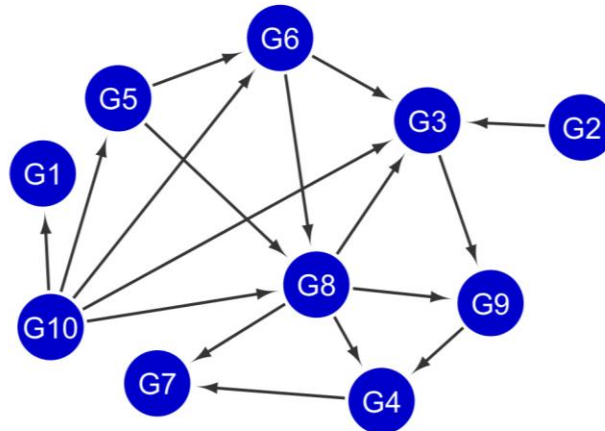


Figure 4.10 DREAM4 size 10 network 5 produced by BANJO tool.
Source: Cytoscape

The Figure 4.10 is of DREAM4 size 10 network 5 produced by BANJO tool. As seen in the diagram, genes 8 and 10 interact with four genes, namely 3, 4, 7, and 9 and 1, 5, 6 and 8 respectively. Genes such as 2, 3, 4, and 9 connect with one gene as clearly seen in the network. Genes numbered 5 and 6 are communicating with two genes.

4.1.3 Causality

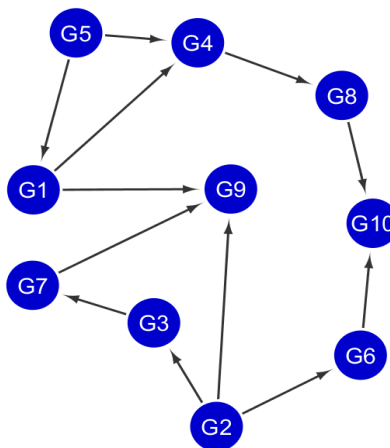


Figure 4.11 Gene inference of DREAM4 size 10 network 1 by causality algorithm.
Source: Cytoscape

As seen in Figure 4.11, all genes are interacting with their corresponding genes. Genes such as 3, 4, 6, 7, and 8 connect with only one gene. Two genes are interacting with two genes, namely 1 with 4 and 9, and number 5 with 1 and 4 as seen in the diagram. Only gene number 2 is interacting with three genes in the network.

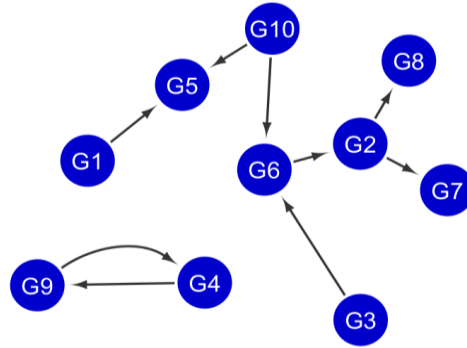


Figure 4.12 Gene inference of DREAM4 size 10 network 2 by causality algorithm.
Source: Cytoscape

As seen in Figure 4.12, there is one bidirected edge between genes 9 and 4. Gene 2 and gene 10 have two interactions with genes. The remaining genes number 1, 3, and 6 have one interface with corresponding genes 5, 6 and 2 respectively as clearly observed from the network diagram.

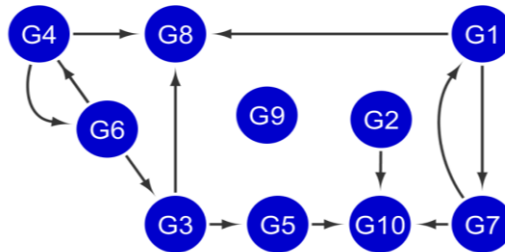


Figure 4.13 Gene inference of DREAM4 size 10 network 3 by causality algorithm.
Source: Cytoscape

As seen in the GRN 4.13, there are two bidirected edges among pairs of two genes namely 4 and 6 and 1 and 7. Gene 9 is a solo gene not communicating with any of the gene in the network. The remaining genes have one interaction. For instance, gene 2 interfaces with gene 10.

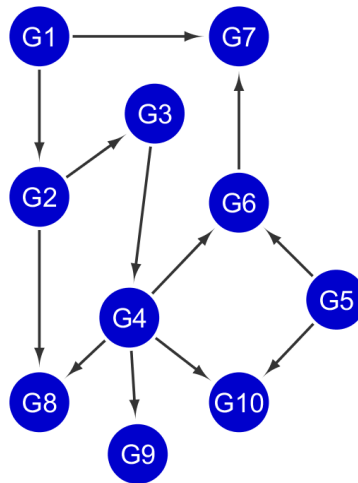


Figure 4.14 Gene inference of DREAM4 size 10 network 4 by causality algorithm.
Source: Cytoscape

In this Figure, there is no bidirected edge. Gene 4 is interacting with four genes, namely 6, 8, 9 and 10. Genes 1, 2 and 5 are communicating with two corresponding genes as seen in Figure 4.14. Last, only one interaction happens between gene 6 and gene 7 as well as gene 3 with gene 4.

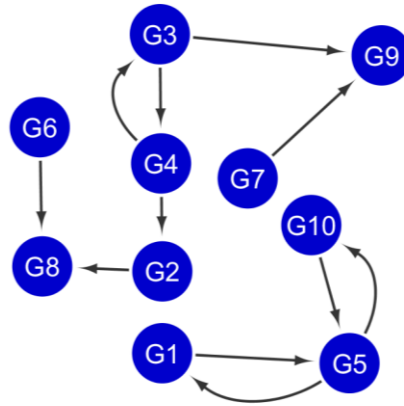


Figure 4.15 Gene inference of DREAM4 size 10 network 5 by causality algorithm.
Source: Cytoscape

This is the Figure of DREAM4 size 10 network 5 by causality algorithm. In this gene network diagram, there are three bidirected edges between genes 1 and 5, 3 and 4 as well as 5 and 10. As clearly observed in the obtained GRN, the remaining gene interacts with only one corresponding gene.

4.2 Comparison of Network Obtained from Steady-State Tools

4.2.1 ARACNe

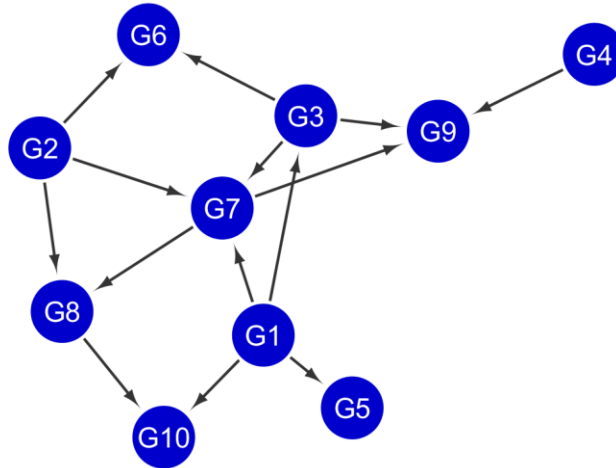


Figure 4.16 DREAM4 size 10 network 1 created by ARACNe using Kendell algorithm.
Source: Cytoscape

As seen in Figure 4.16, gene 7 interacts with two genes, namely 8 and 9. Also, gene 1 communicates with four genes, 3, 5, 7 and 10, as observed in the diagram. Genes namely 4 and 8 interface with only one gene such 9 and 10 respectively. Lastly, gene 3 is interacting with three genes, namely 6, 7 and 9. Similarly gene 2 communicates with 6, 7 and 8.

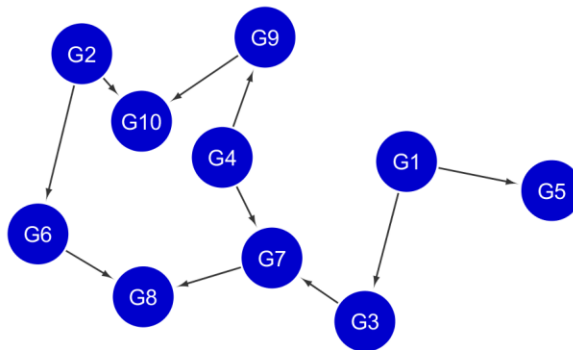


Figure 4.17 DREAM4 size 10 network 1 created by ARACNe using Pearson algorithm.
Source: Cytoscape

This is the DREAM4 size 10 network 1 created by ARACNe using the Pearson algorithm. In this gene network diagram, there are single gene to gene interactions. For instance, genes 6 and 7 both communicate with gene 8. Gene 1 interfaces with genes 5 and 3, and gene 2 with 6 and 10 as clearly observed in the obtained GRN.

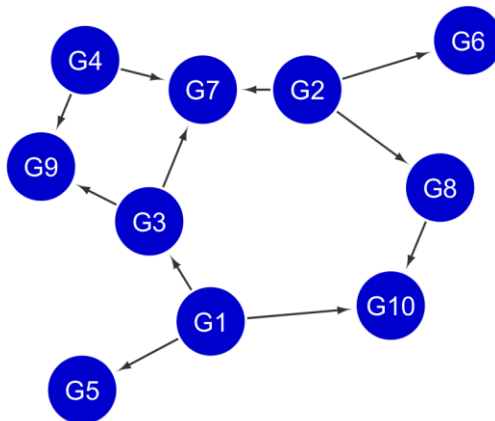


Figure 4.18 DREAM4 size 10 network 1 created by ARACNe using Spearman algorithm.
Source: Cytoscape

As seen in the GRN 4.18, there are two genes interacting with two corresponding genes, namely genes 3 and 4 both with 7 and 9. There is only one interaction between

gene 8 and gene 10. The remaining genes have three interfaces with genes. For instance, gene 1 communicates with 3, 5 and 10.

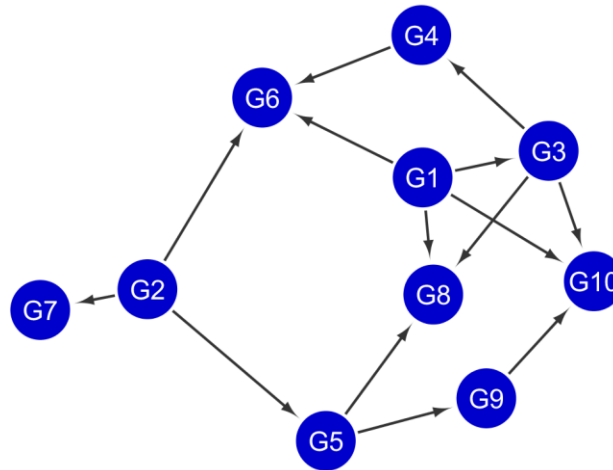


Figure 4.19 DREAM4 size 10 network 2 created by ARACNe using Kendell algorithm.
Source: Cytoscape

In this Figure 4.19, gene 1 interacts with four genes, namely 3, 6, 8 and 10. There are two genes that are interacting with three genes. For instance, gene 2 communicates with genes 4, 8 and 10. As clearly observed in the network, gene 5 interfaces with two genes numbered 8 and 9. There is one gene to gene interaction which is between genes 9 and 10.

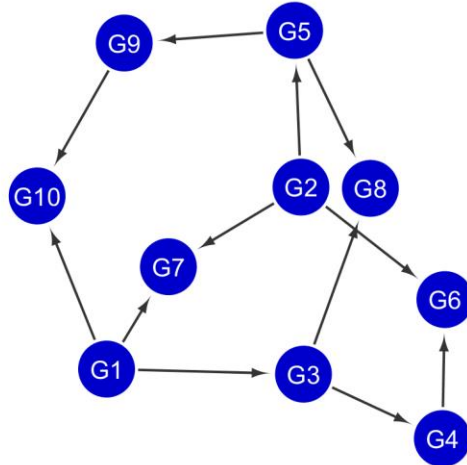


Figure 4.20 DREAM4 size 10 network 2 created by ARACNe using Pearson algorithm.
Source: Cytoscape

As seen in the GRN 4.20, there are two genes interacting with two corresponding genes. For example, gene 5 interfaces with genes 8 and 9. There are two genes which have one gene to gene interaction, such as gene 9 communicates with 10. The remaining genes have three connections with genes. For instance, gene 1 communicates with 3, 7 and 10.

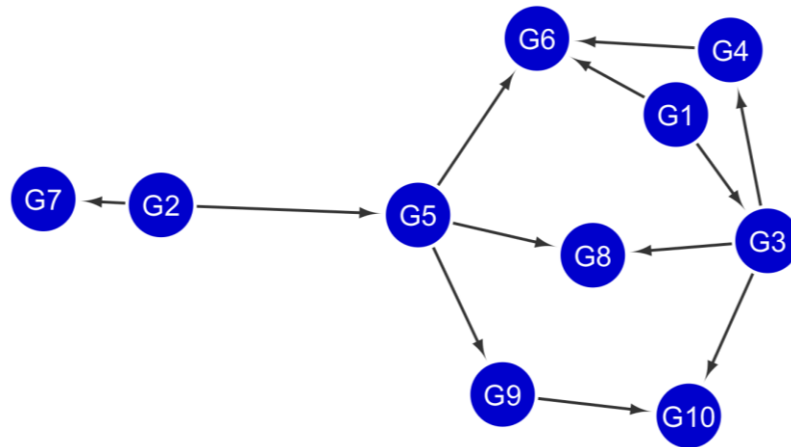


Figure 4.21 DREAM4 size 10 network 2 created by ARACNe using Spearman algorithm.
Source: Cytoscape

This Figure 4.21 is of DREAM4 size 10 network 2 created by ARACNe using Spearman algorithm. In this gene network diagram, there is a single gene to gene

interaction. For instance, gene 9 interfaces with 10. As seen in the Figure 4.21, gene 2 connects with two genes namely 5 and 7. Gene 5 communicates with genes 6, 8 and 9 as clearly observed in the obtained GRN.

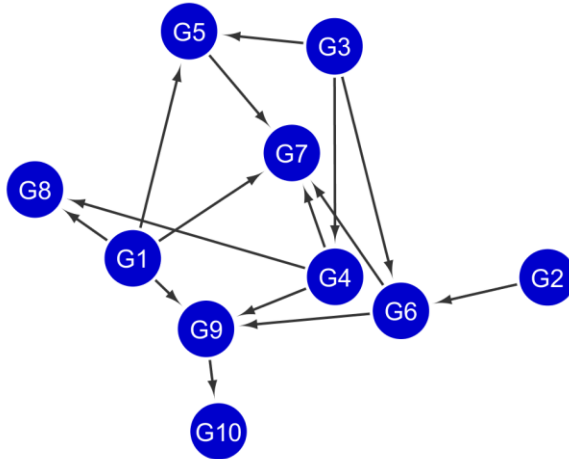


Figure 4.22 DREAM4 size 10 network 3 created by ARACNe using Kendall algorithm.
Source: Cytoscape

As seen in Figure 4.22, gene 1 interacts with four genes, namely 5, 7, 8 and 9. Also, gene 3 communicates with three genes 4, 5, and 6 as observed in the diagram. Gene 4 interfaces with 7 and 9. Lastly, there are three genes which have one to one gene interactions.

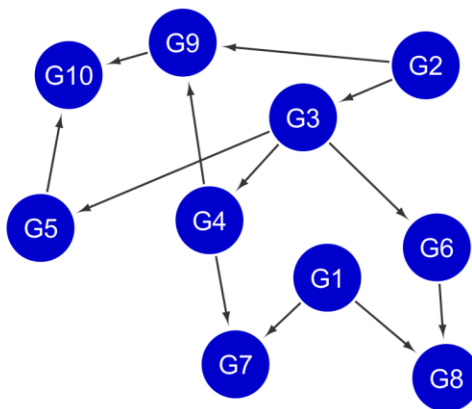


Figure 4.23 DREAM4 size 10 network 3 created by ARACNe using Pearson algorithm.
Source: Cytoscape

As seen in the GRN 4.23, gene 4 interacts with two genes, namely 7 and 9. There is double communication between two genes as seen in the obtained network. For example, gene 1 interfaces with 7 and 8. Similarly gene 2 interacts with 3 and 9. Gene 3 has three connections with genes, namely genes numbered 4, 5 and 6.

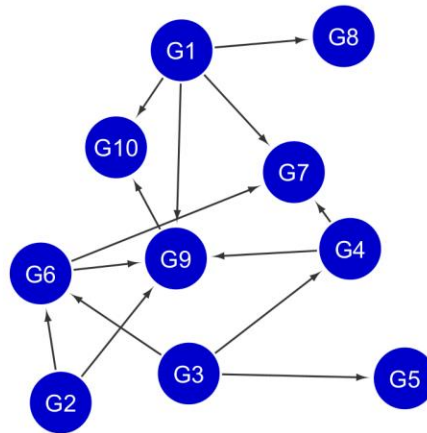


Figure 4.24 DREAM4 size 10 network 3 created by ARACNe using Spearman algorithm.
Source: Cytoscape

This Figure 4.24 is of DREAM4 size 10 network 3 created by ARACNe using the Spearman algorithm. In this gene network diagram, there is a single gene to gene interaction. For instance, gene 9 interfaces with 10. As seen in Figure 4.24, gene 2 connects with two genes namely 7 and 9. Genes 4 and 6 both interact with the same two genes, namely 7 and 9 as clearly observed in the obtained GRN. Lastly, gene 3 communicates with genes 4, 5 and 6.

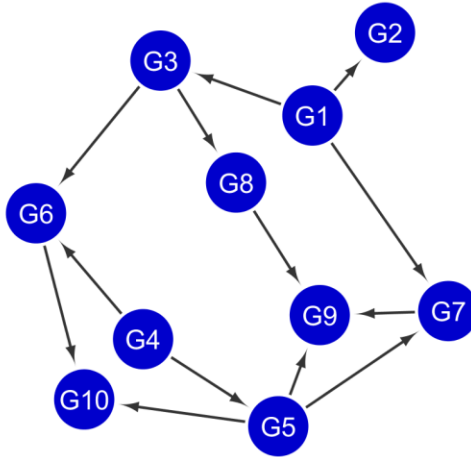


Figure 4.25 DREAM4 size 10 network 4 created by ARACNe using Kendall algorithm.
Source: Cytoscape

As seen in Figure 4.25, gene 1 interacts with three genes, namely 2, 3 and 7. Also, gene 3 communicates with two genes 6 and 8 as observed in the diagram. Gene 4 interfaces with 5 and 6. Lastly, there are two genes which have one to one gene interaction. For example, gene 6 communicates with 10 and gene 7 with 9.

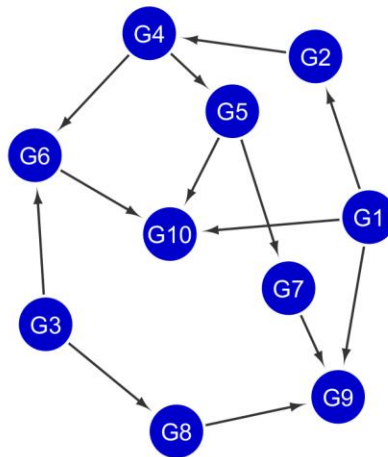


Figure 4.26 DREAM4 size 10 network 4 created by ARACNe using Pearson algorithm.
Source: Cytoscape

As seen in the GRN 4.26, gene 4 interacts with two genes, namely 6 and 5. There is a gene 1 which has three interactions with genes 2, 9 and 10 as seen in the obtained

network. Gene 3 interfaces with two genes, namely gene 6 and 8. There are two single gene to gene interactions, namely gene 8 with 9 as well as gene 2 with 4.

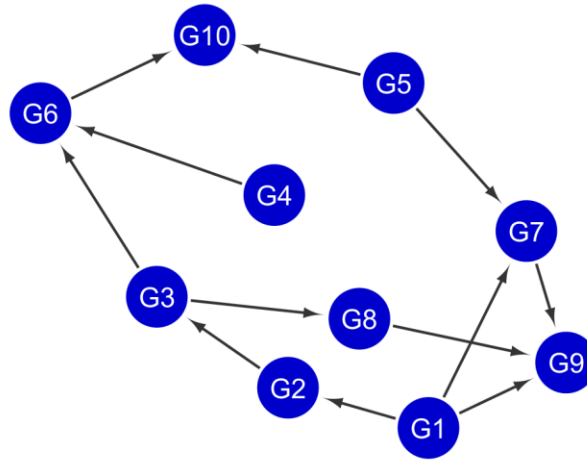


Figure 4.27 DREAM4 size 10 network 4 created by ARACNe using Spearman algorithm.
Source: Cytoscape

This Figure 4.27 is of DREAM4 size 10 network 4 was created by the ARACNe using the Spearman algorithm. In this gene network diagram, there is a single gene to gene interaction. For instance, gene 4 interfaces with 6. As seen in Figure 4.27, gene 3 connects with two genes, namely 6 and 8. Gene 5 and 6 both interact with the same gene, namely 10 as clearly observed in the obtained GRN. Last, gene 1 communicates with genes 2, 7 and 9.

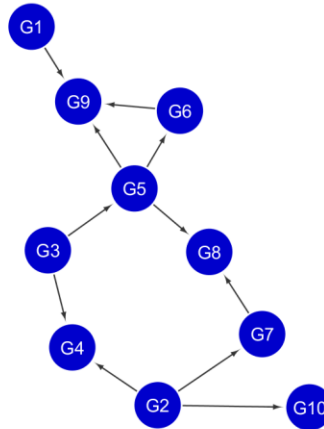


Figure 4.28 DREAM4 size 10 network 5 created by ARACNe using Kendall algorithm.
Source: Cytoscape

As seen in Figure 4.28, gene 5 interacts with three genes, namely 6, 8 and 9. Also, gene 3 communicates with two genes 4 and 5 as observed in the diagram. Gene 7 interfaces with 8 which is a single gene to gene interaction. Last, genes such as 1 and 6 both communicate with gene 9.

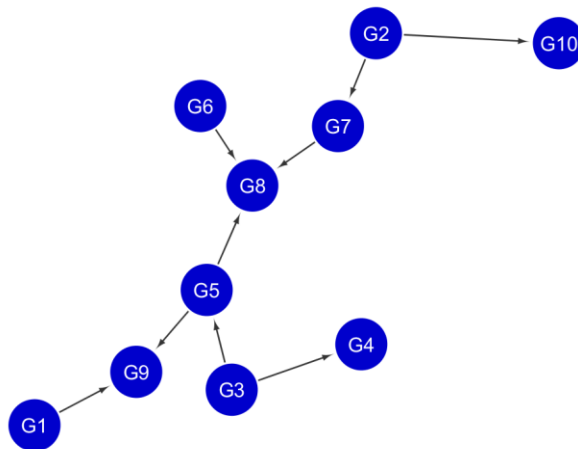


Figure 4.29 DREAM4 size 10 network 5 created by ARACNe using Pearson algorithm.
Source: Cytoscape

As seen in the GRN 4.29, gene 2 interacts with two genes, namely 7 and 10. There is a one common interaction with gene 8 by genes 6 and 7 as seen in the obtained

network. Gene 3 interfaces with two genes, namely gene 4 and 5. Last, there is a single gene to gene interaction between gene 1 and 9.

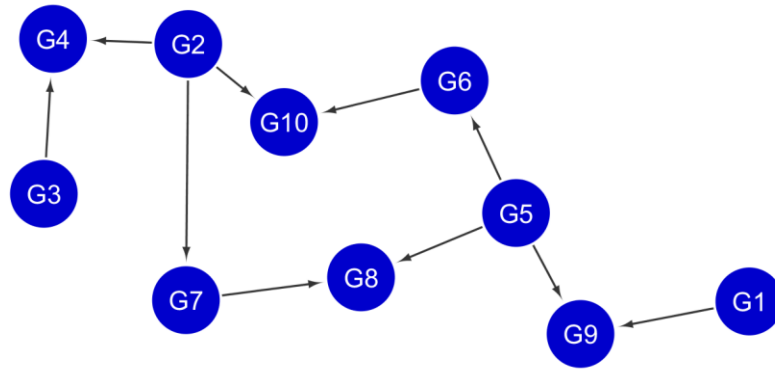


Figure 4.30 DREAM4 size 10 network 5 created by ARACNe using Spearman algorithm.

Source: Cytoscape

This Figure 4.30 is of DREAM4 size 10 network 5 was created by the ARACNe using the Spearman algorithm. In this gene network diagram, there are many gene to gene interactions. For instance, gene 3 interfaces with 4. As seen in Figure 4.30, gene 2 communicates with genes such as 4, 7 and 10.

4.2.2 GENIE3

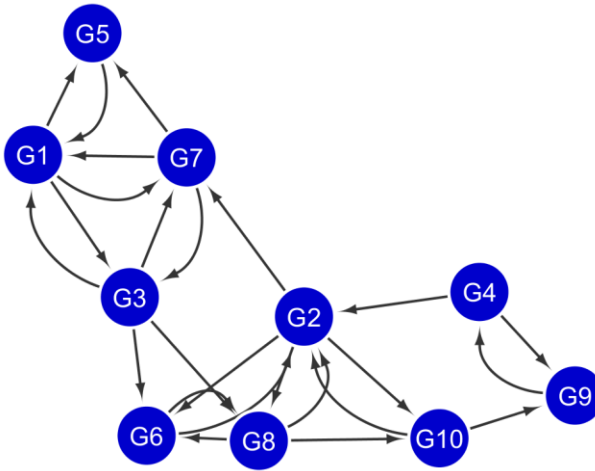


Figure 4.31 DREAM4 size 10 network 1 built by GENIE3.

Source: Cytoscape

As seen in the GRN 4.31, there are many bidirected edges among pairs of two genes. For example, gene 2 has bidirected connections with genes 6, 8 and 10. Gene 3 interacts with two genes such as 6 and 8. Last, there are two single gene to gene interactions between gene 8 and 10 and gene 10 communicates with gene 9 as clearly observed in the obtained gene network.

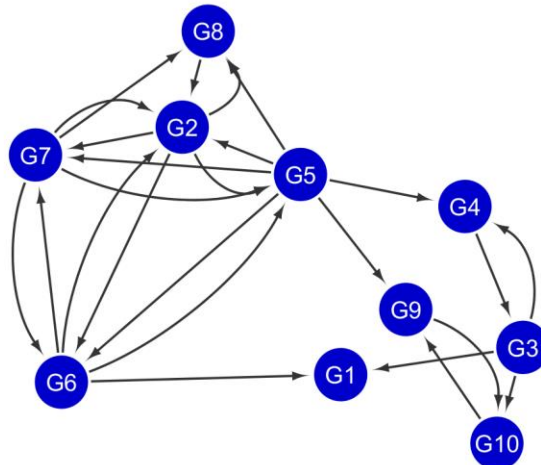


Figure 4.32 DREAM4 size 10 network 2 built by GENIE3.

Source: Cytoscape

This Figure 4.30 is of DREAM4 size 10 network 2 was built by GENIE3. As seen in the GRN 4.32, there are many bidirected edges among pairs of two genes. For example, gene 2 has bidirected connections with genes 5, 6, 7 and 8. Gene 3 interacts with two genes such as 1 and 10. Last, there is only one gene to gene interaction between gene 6 and 1 as clearly observed in the obtained gene network.

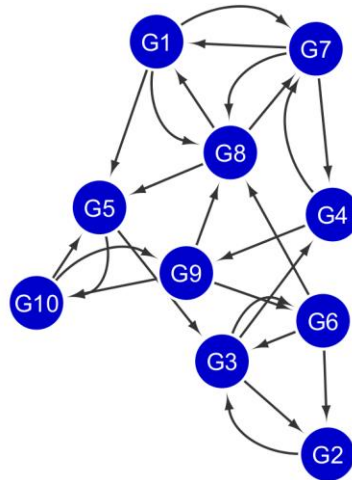


Figure 4.33 DREAM4 size 10 network 3 built by GENIE3.
Source: Cytoscape

As seen in the GRN 4.33, there are many bidirected edges among pairs of two genes. For example, gene 7 has bidirected connections with genes 4 and 8. Gene 9 interacts with two genes such as 6 and 8. Last, there are many single gene to gene interactions as seen in the obtained diagram. For instance, there is a connection between gene 8 and 5 and gene 5 communicates with gene 3 as clearly observed in Figure.

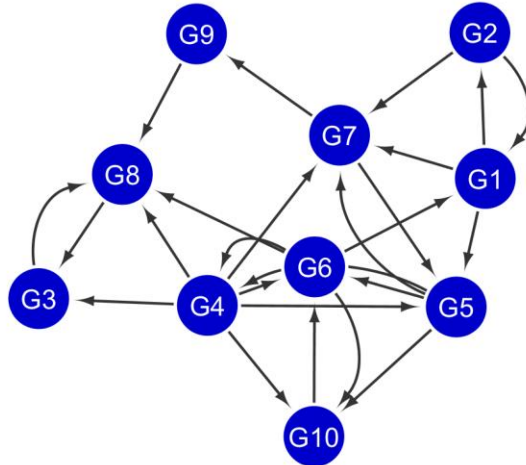


Figure 4.34 DREAM4 size 10 network 4 built by GENIE3.
Source: Cytoscape

As seen in the gene network 4.34, there are many bidirected edges among pairs of two genes. For example, gene 1 has bidirected connections with gene 2. Gene 6 interacts with two genes such as 1 and 8. As seen in Figure, gene 4 interfaces with four genes number 3, 7, 8 and 10. Last, there are two single gene to gene interactions between gene 2 and 7 and gene 7 communicates with gene 9 as clearly observed in the obtained gene network.

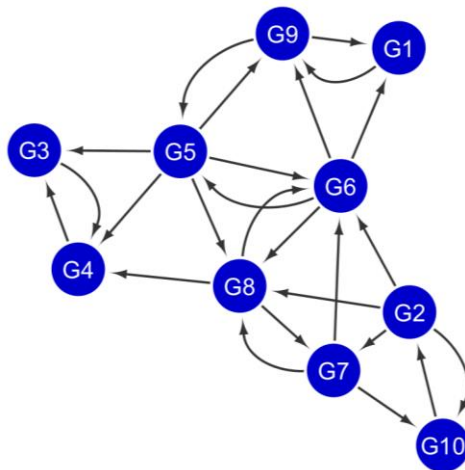


Figure 4.35 DREAM4 size 10 network 5 built by GENIE3.
Source: Cytoscape

This Figure 4.35 is of DREAM4 size 10 network 5 was built by GENIE3. As seen in the GRN 4.35, there are many bidirected edges among pairs of two genes. For example, gene 4 has bidirected connections with gene 3. Gene 6 interacts with two genes such as 5 and 1. Gene 5 interfaces with three genes, namely 3, 4 and 8. Lastly, there is one gene to gene connection between gene 8 and 4 as clearly observed in the obtained gene network.

4.2.3 CLR

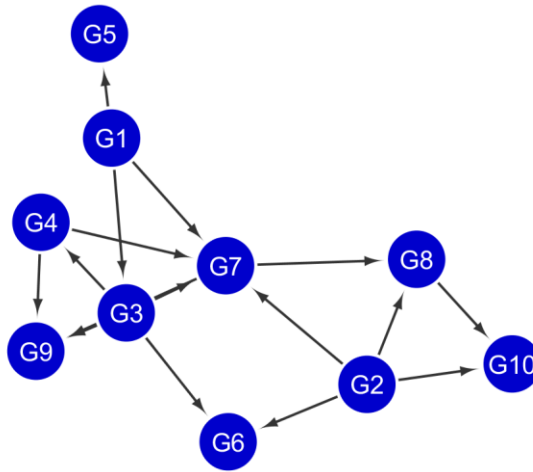


Figure 4.36 DREAM4 size 10 network 1 generated by CLR using Kendell algorithm.
Source: Cytoscape

This Figure 4.36 is of DREAM4 size 10 network 1 was generated by the CLR using the Kendell algorithm. In this gene network diagram, there are two gene to gene interactions. For instance, gene 7 interfaces with 8 and gene 8 with 10. As seen in Figure 4.36, gene 3 connects with two genes, namely 4 and 6. Gene 2 communicates with four genes, namely 6, 7, 8 and 10 as clearly observed in the obtained GRN. Last, gene 1 communicates with genes 3, 5 and 7.

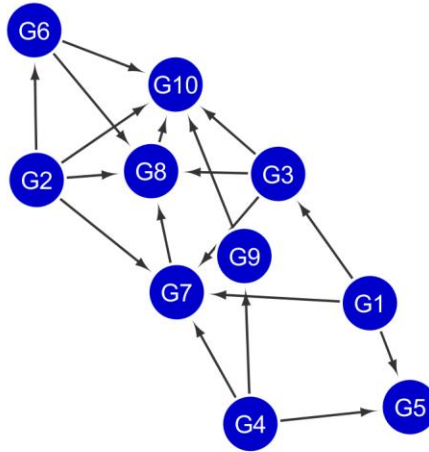


Figure 4.37 DREAM4 size 10 network 1 generated by CLR using Pearson algorithm.
Source: Cytoscape

As seen in Figure 4.37, gene 1 interacts with three genes, namely 3, 5 and 7. Also, gene 6 communicates with two genes 8 and 10 as observed in the diagram. Gene 2 has four connections with 4, 7, 8 and 10. Last, genes such as 9 and 8 both communicate with gene 10 which is a single gene to gene interaction.

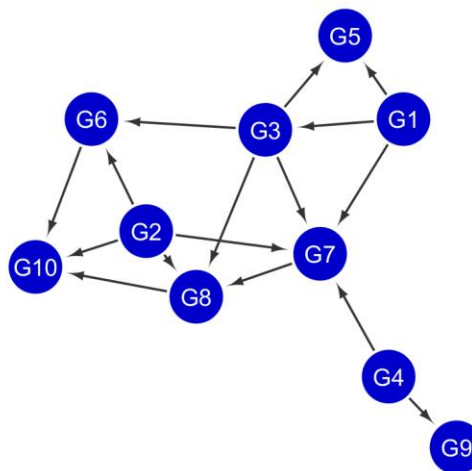


Figure 4.38 DREAM4 size 10 network 1 generated by CLR using Spearman algorithm.
Source: Cytoscape

This Figure 4.38 is of DREAM4 size 10 network 1 was generated by the CLR using the Spearman algorithm. In this gene network diagram, there are many gene to gene interactions. For instance, gene 7 interfaces with 8. Gene 4 connects with two genes,

namely 7 and 9. As seen in Figure 4.38, gene 1 communicates with genes 3, 7 and 5. Last, gene 3 interfaces with four genes such as 5, 6, 7 and 8 as clearly seen in the obtained gene network.

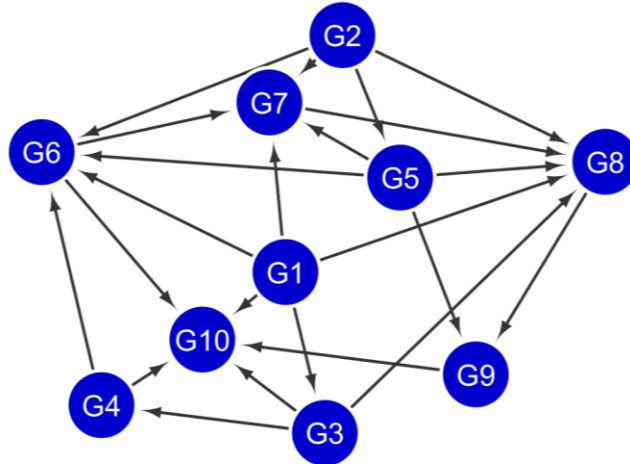


Figure 4.39 DREAM4 size 10 network 2 generated by CLR using Kendall algorithm.
Source: Cytoscape

In this gene network diagram, there is no single gene to gene interaction. As seen in Figure 4.39, there are five genes, namely 1, 2, 3, 5 and 7 which are interacting with gene 8. Gene 6 interfaces with two genes, namely 7 and 10. Last, gene 2 communicates with genes 6, 5 and 7.

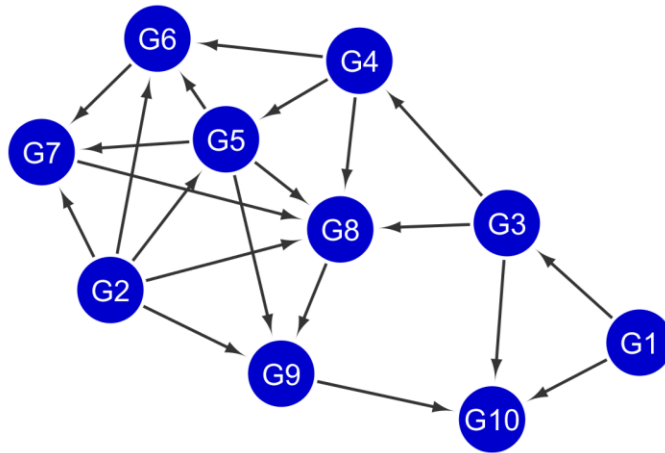


Figure 4.40 DREAM4 size 10 network 2 generated by CLR using Pearson algorithm.
Source: Cytoscape

As seen in Figure 4.40, gene 3 interacts with three genes, namely 4, 8 and 10. Also, gene 1 communicates with two genes 3 and 10 as observed in the diagram. Gene 5 has four connections with genes 6, 7, 8 and 9. As observed in the network, gene 9 is connected to gene 10 and 8 with gene 9. Last, gene 2 interacts with genes number 5, 6, 7, 8 and 9.

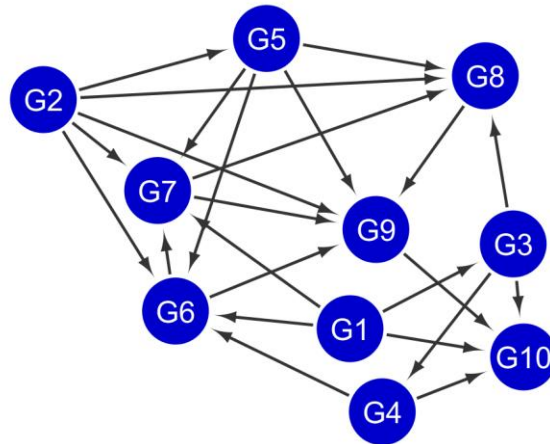


Figure 4.41 DREAM4 size 10 network 2 generated by CLR Spearman algorithm.
Source: Cytoscape

This Figure 4.41 is of DREAM4 size 10 network 2 was generated by the CLR the Spearman algorithm. In this gene network diagram, there are two gene to gene interactions. For instance, gene 9 interfaces with 10. Gene 4 connects with two genes, namely 6 and 10. As seen in Figure 4.41, gene 5 communicates with genes 6, 7, 8 and 9. Last, gene 2 interfaces with five genes such as 5, 6, 7, 8 and 9 as clearly seen in the obtained gene network.

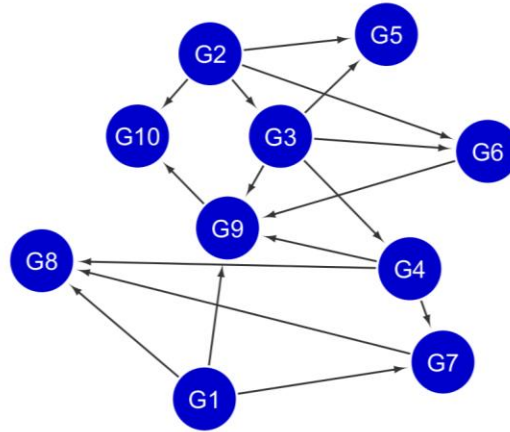


Figure 4.42 DREAM4 size 10 network 3 generated by CLR using Kendall algorithm.
Source: Cytoscape

In this gene network diagram, there are two single gene to gene interactions between gene 9 and 10 and gene 6 and 9. As seen in Figure 4.42, gene 1 interacts with genes 7, 8 and 9. Gene 4 interfaces with two genes, namely 7 and 9. Last, gene 2 communicates with four genes, namely 3, 6, 5 and 10.

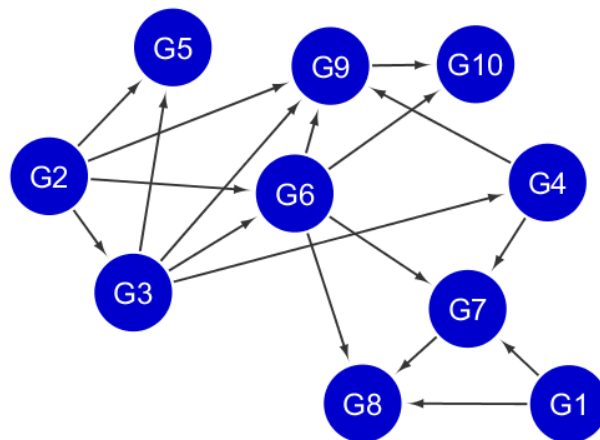


Figure 4.43 DREAM4 size 10 network 3 generated by CLR using Pearson algorithm.
Source: Cytoscape

This Figure 4.43 is of DREAM4 size 10 network 3 was generated by the CLR using the Pearson algorithm. As seen in Figure 4.43, there are two single gene to gene interactions between genes 9 and 10 as well as genes 7 and 8. Also, gene 1 communicates

with two genes 7 and 8 as observed in the diagram. Gene 2 has four connections with genes 3, 5, 6 and 9.

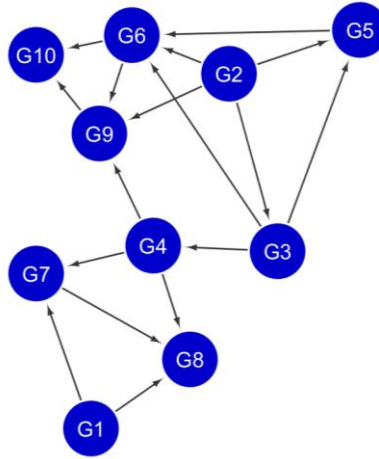


Figure 4.44 DREAM4 size 10 network 3 generated by CLR using Spearman algorithm.
Source: Cytoscape

In this gene network diagram 4.44, there are three gene to gene interactions. For instance, gene 5 interfaces with gene 6, gene 7 with 8 and gene 9 with gene 10. Gene 4 connects with three genes, namely 7, 8 and 9. As seen in Figure 4.44, gene 1 communicates with two corresponding genes namely 7 and 8.

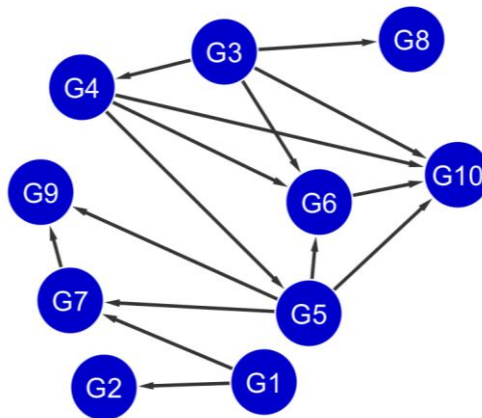


Figure 4.45 DREAM4 size 10 network 4 generated by CLR using Kendell algorithm.
Source: Cytoscape

This Figure 4.45 is of DREAM4 size 10 network 4 was generated by the CLR using the Kendell algorithm. In this gene network diagram, there are two gene to gene interactions. For instance, gene 6 interfaces with 10. Gene 4 connects with three genes, namely 4, 6 and 10. As seen in Figure 4.45, gene 5 communicates with genes 6, 7, 10 and 9. Last, gene 1 interfaces with two genes such as 2 and 7 as clearly seen in the obtained gene network.

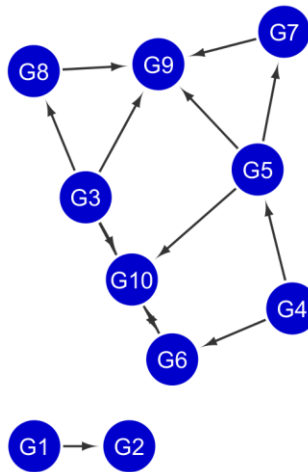


Figure 4.46 DREAM4 size 10 network 4 generated by CLR using Pearson algorithm.
Source: Cytoscape

This Figure 4.46 is of DREAM4 size 10 network 4 was generated by the CLR using the Pearson algorithm. As seen in Figure 4.46, there are two single gene to gene interactions between gene 1 and 2 and gene 6 and 10. Also, genes 8 and 7 both communicate with gene 9 as observed in the diagram. Gene 4 interfaces with genes 5 and 6. Last, gene 3 has three connections with genes 8, 9 and 10.

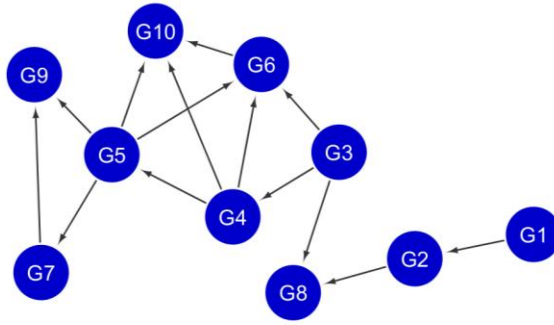


Figure 4.47 DREAM4 size 10 network 4 generated by CLR using Spearman algorithm.
Source: Cytoscape

In this gene network diagram 4.47, there are four gene to gene interactions. For instance, gene 1 interfaces with gene 2, gene 2 with 8, gene 7 with 9 and finally gene 6 with gene 10. Gene 4 connects with three genes, namely 5, 6 and 10. As seen in Figure 4.47, gene 5 communicates with four corresponding genes namely 6, 7, 9 and 10.

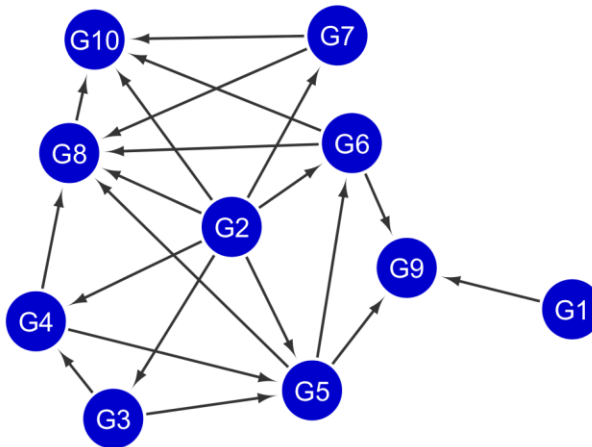


Figure 4.48 DREAM4 size 10 network 5 generated by CLR using Kendall algorithm.
Source: Cytoscape

This Figure 4.48 is of DREAM4 size 10 network 5 was generated by the CLR using the Kendall algorithm. In this gene network diagram, there are two gene to gene interactions. For instance, gene 8 interfaces with 10. Gene 4 connects with two genes, namely 5 and 8. As seen in Figure 4.48, gene 5 communicates with three genes 6, 8, and

9. Last, gene 2 interfaces with seven genes such as 3, 4, 5, 7, 8 and 10 as clearly seen in the obtained gene network.

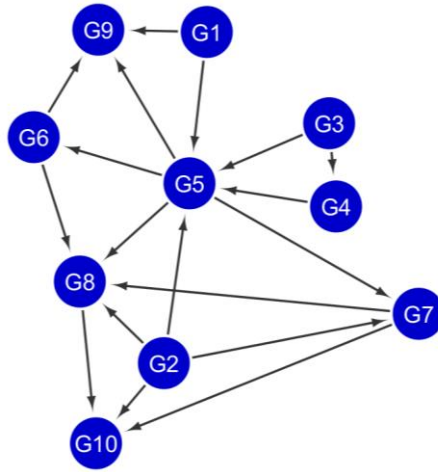


Figure 4.49 DREAM4 size 10 network 5 generated by CLR using Pearson algorithm.
Source: Cytoscape

As seen in Figure 4.49, gene 3 interacts with two genes, namely 4, and 5. Also, gene 1 communicates with two genes 5 and 9 as observed in the diagram. Gene 5 has four connections with genes 6, 7, 8 and 9. Last, there are two gene to gene connections between gene 4 and 5 as well as between gene 8 and 10.

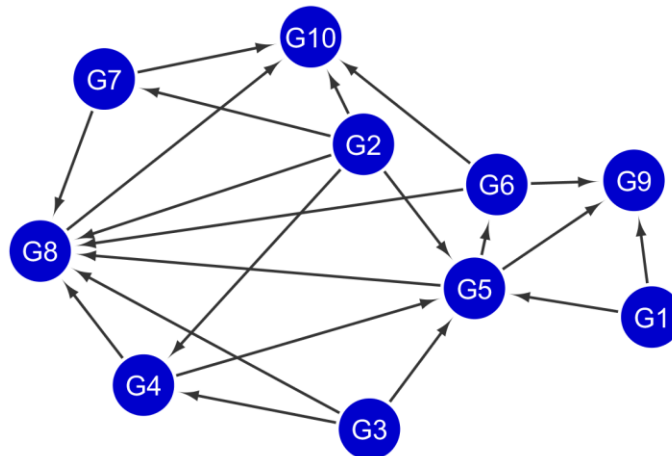


Figure 4.50 DREAM4 size 10 network 5 generated by CLR using Spearman algorithm.
Source: Cytoscape

This Figure 4.50 is of DREAM4 size 10 network 5 was generated by the CLR using the Spearman algorithm. In this gene network diagram, there is no gene to gene interaction. Gene 4 connects with two genes, namely 5 and 8. As seen in Figure 4.50, gene 5 communicates with three genes 6, 8, and 9. Last, gene 2 interfaces with seven genes such as 3, 4, 5, 7, 8 and 10 as clearly seen in the obtained gene network.

4.2.4 MRNET

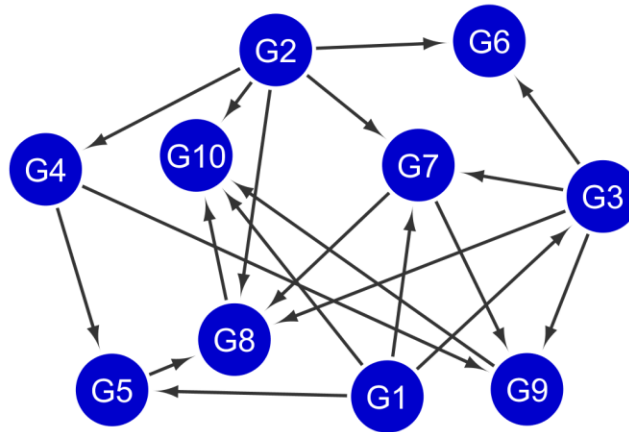


Figure 4.51 Gene network of DREAM4 size 10 network 1 using Kendell method by MRNET.

Source: Cytoscape

As seen in Figure 4.51, gene 3 interacts with four genes, namely 6, 7, 8 and 9. Also, gene 1 communicates with four genes 3, 5, 7 and 10 as observed in the diagram. Gene 2 has five connections with genes 4, 6, 7, 8 and 10. Last, there are four gene to gene connections in this gene network. For example, genes 8 and 9 both interact with only gene number 10.

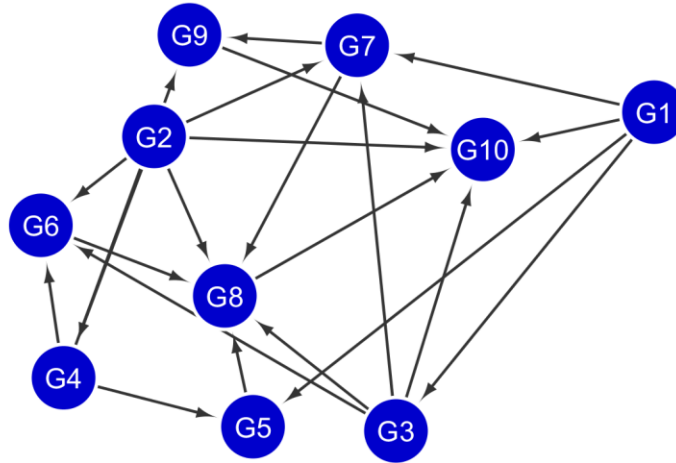


Figure 4.52 Gene network of DREAM4 size 10 network 1 using Pearson method by MRNET.

Source: Cytoscape

This Figure 4.52 is Gene network of DREAM4 size 10 network 1 using the Pearson method by the MRNET. As seen in Figure 4.52, gene 1 interacts with four genes, namely 3, 5, 7 and 10. Also, gene 3 has four connections with genes 6, 7, 8 and 10. As observed in the network, gene 7 interfaces with genes 9 and 8. Last, there are two gene to gene connections in this gene network. For example, gene number 6 only connects with gene 8 as observed in the obtained gene network diagram 4.52.

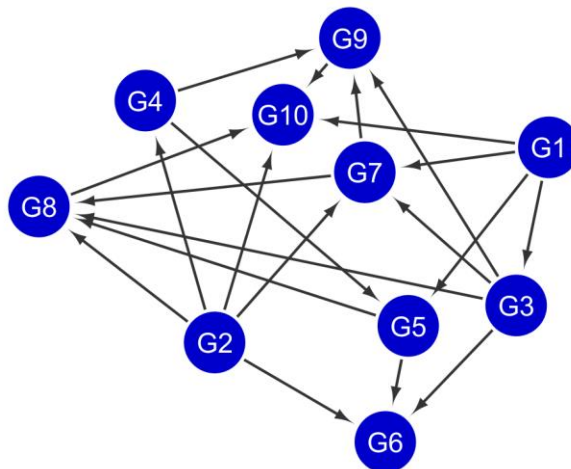


Figure 4.53 Gene network of DREAM4 size 10 network 1 using Spearman method by MRNET.

Source: Cytoscape

This Figure 4.53 is gene network of DREAM4 size 10 network 1 using the Spearman method by the MRNET. In this gene network diagram, there are three gene to gene interactions. For instance, genes 8 and 9 both communicate with gene number 10. Gene 4 connects with two genes, namely 5 and 9. Also, as observed in the network, gene 7 interfaces with genes 9 and 8. As seen in Figure 4.53, gene 3 communicates with four genes 6, 7, 8, and 9. Last, gene 2 interfaces with five genes such as 4, 7, 6, 8 and 10 as clearly seen in the obtained gene network.

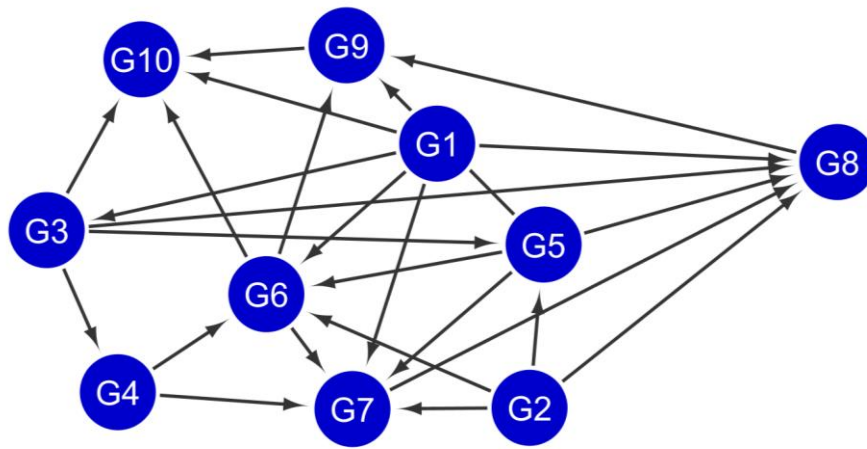


Figure 4.54 Gene network of DREAM4 size 10 network 2 using Kendall method by MRNET.
Source: Cytoscape

This Figure 4.54 is the gene network of DREAM4 size 10 network 2 using the Kendall method by the MRNET. In this gene network diagram, there are two gene to gene interactions. For instance, gene 8 interfaces with 9 and gene 9 interact with gene 10. Gene 4 connects with two genes, namely 6 and 7. As seen in the Figure 4.54, gene 5 communicates with four genes 6, 7, 8, and 9. Last, gene 1 interfaces with three genes such as 6, 7 and 9 as clearly seen in the obtained gene network.

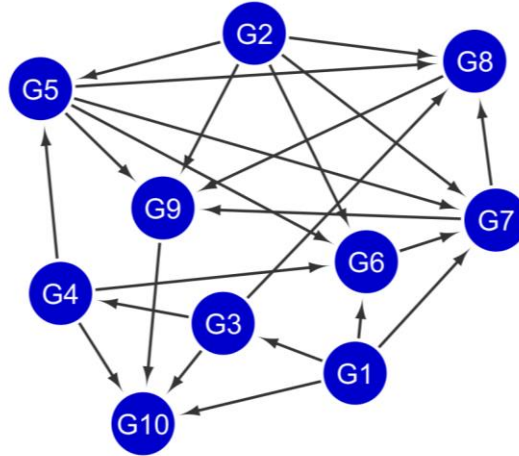


Figure 4.55 Gene network of DREAM4 size 10 network 2 using Pearson method by MRNET.

Source: Cytoscape

In this gene network diagram 4.55, there are four gene to gene interactions. For instance, gene 9 communicates with gene number 10. Gene 4 connects with three genes, namely 5, 6 and 10. As seen in Figure 4.55, gene 5 communicates with four genes 6, 7, 8, and 9. Last, gene 2 interfaces with five genes such as 5, 7, 6, 8 and 9 as clearly seen in the obtained gene network.

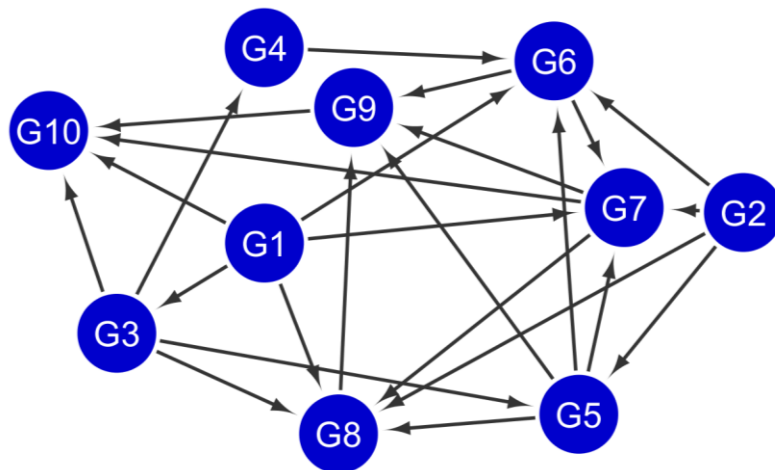


Figure 4.56 Gene network of DREAM4 size 10 network 2 using Spearman method by MRNET.

Source: Cytoscape

This Figure 4.56 is gene network of DREAM4 size 10 network 2 using the Spearman method by the MRNET. In this gene network diagram, there are three gene to gene interactions. For instance, genes 8 and 9 both communicate with gene number 10. Gene 6 connects with two genes, namely 7 and 9. Also, as observed in the network, gene 7 interfaces with three genes 10, 9 and 8. As seen in Figure 4.56, gene 3 communicates with four genes 4, 5, 8, and 10. Last, gene 1 interfaces with five genes such as 3, 7, 6, 8 and 10 as clearly seen in the obtained gene network.

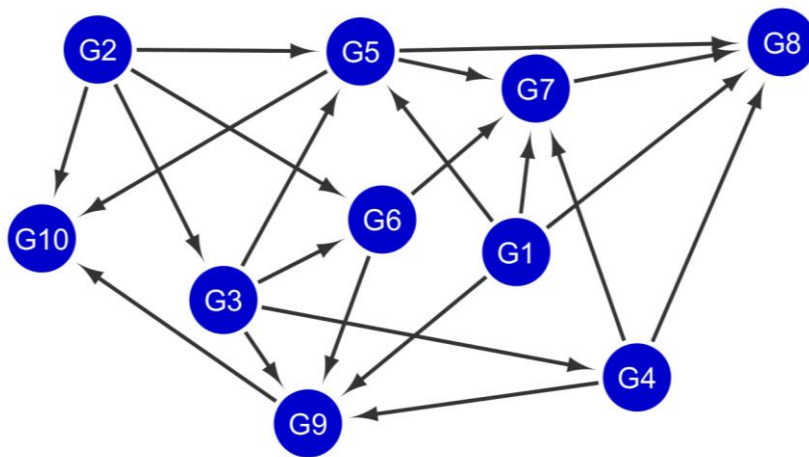


Figure 4.57 Gene network of DREAM4 size 10 network 3 using Kendell method by MRNET.

Source: Cytoscape

In this gene network diagram, there are two gene to gene interactions. For instance, gene 9 interfaces with 10 and gene 7 interact with gene 8. Gene 4 connects with three genes, namely 9, 8 and 7. As seen in Figure 4.57, gene 1 communicates with four genes 5, 7, 8, and 9. Last, gene 6 interfaces with two genes such as 7 and 9 as clearly seen in the obtained gene network.

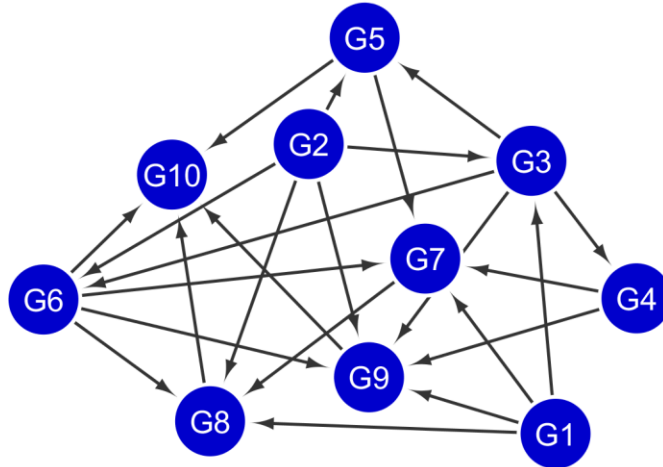


Figure 4.58 Gene network of DREAM4 size 10 network 3 using Pearson method by MRNET.

Source: Cytoscape

This Figure 4.58 is gene network of DREAM4 size 10 network 3 using the Pearson method by the MRNET. As seen in figure 4.58, gene 1 interacts with five genes, namely 3, 8, 7, 9 and 10. Also, gene 3 has four connections with genes 6, 9, 5 and 4. As observed in the network, gene 6 interfaces with two genes 7 and 10. Last, there are three gene to gene connections in this gene network. For example, gene number 7 connects with gene 8. In addition, genes 8 and 9 both connect to gene 10 as seen in the gene network.

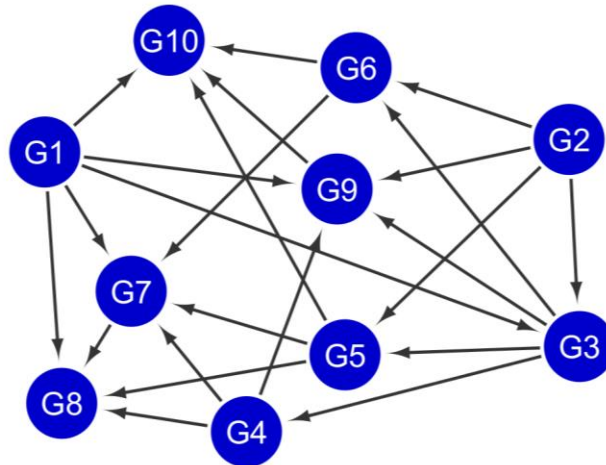


Figure 4.59 Gene network of DREAM4 size 10 network 3 using Spearman method by MRNET.

Source: Cytoscape

This Figure 4.59 is gene network of DREAM4 size 10 network 3 using the Spearman method by the MRNET. In this gene network diagram, there are two gene to gene interactions. For instance, gene 9 communicates with gene number 10. Gene 6 connects with two genes, namely 7 and 10. Also, as observed in the network, gene 5 interfaces with three genes 10, 7 and 8. As seen in Figure 4.59, gene 3 communicates with four genes 10, 7 and 8. As seen in Figure 4.59, gene 3 communicates with four genes 4, 5, 6, and 9. Last, gene 1 interfaces with five genes such as 3, 7, 9, 8 and 10 as clearly seen in the obtained gene network.

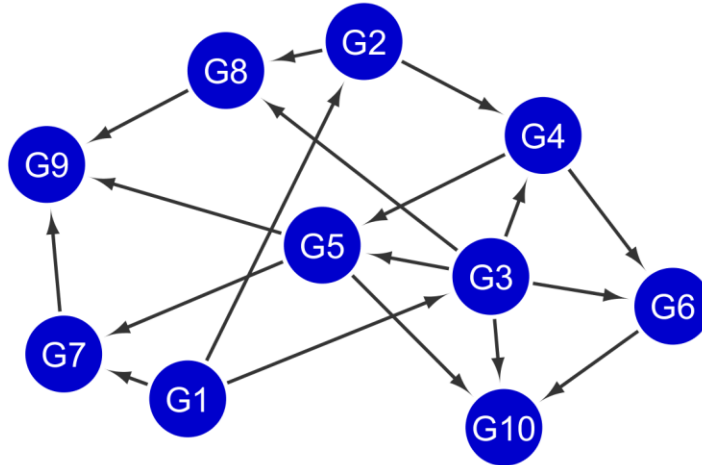


Figure 4.60 Gene network of DREAM4 size 10 network 4 using Kendall method by MRNET.

Source: Cytoscape

In this gene network diagram, there are three gene to gene interactions. For instance, genes 8 and 7 both interface with gene number 9. Gene 1 connects with three genes, namely 3, 2 and 7. As seen in Figure 4.60, gene 2 communicates with two genes 5 and 6. Last, gene 3 interfaces with five genes such as 4, 5, 10, 6 and 8 as clearly seen in the obtained gene network.

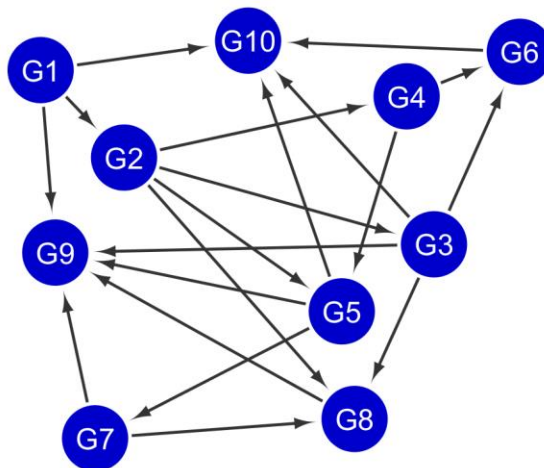


Figure 4.61 Gene network of DREAM4 size 10 network 4 using Pearson method by MRNET.

Source: Cytoscape

This Figure 4.61 is gene network of DREAM4 size 10 network 4 using the Pearson method by the MRNET. As seen in Figure 4.61, gene 1 interacts with three genes, namely 2, 9 and 10. Also, gene 3 has four connections with genes 6, 9, 8, and 10. As observed in the network, gene 7 interfaces with genes 8 and 9. Last, there are three gene to gene connections in this gene network. For example, gene number 8 connects with gene 9. In addition, genes 6 and 5 both connect to gene 10 as seen in the gene network.

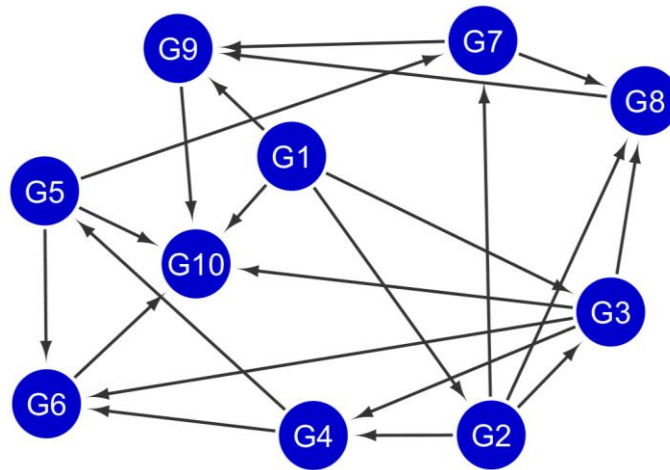


Figure 4.62 Gene network of DREAM4 size 10 network 4 using Spearman method by MRNET.
Source: Cytoscape

This Figure 4.62 is gene network of DREAM4 size 10 network 4 using the Spearman method by the MRNET. In this gene network diagram, there are four gene to gene interactions. For instance, genes 9 and 6 both communicate with gene number 10. Gene 4 connects with two genes, namely 5 and 6. Also, as observed in the network, gene 5 interfaces with three genes 10, 7 and 6. Last, as seen in Figure 4.62, gene 2 interfaces with four genes 4, 3, 8, and 7.

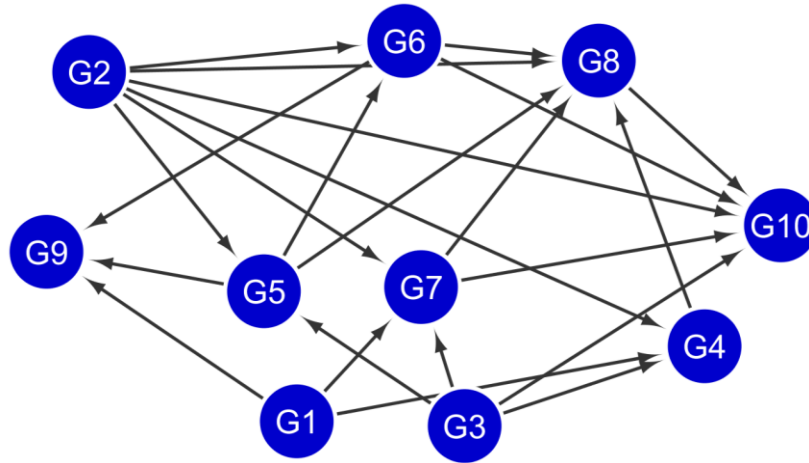


Figure 4.63 Gene network of DREAM4 size 10 network 5 using Kendall method by MRNET.

Source: Cytoscape

In this gene network diagram, there are two gene to gene interactions. For instance, gene 8 interfaces with gene number 10 and gene 4 connects with gene 8. Gene 1 communicates with three genes, namely 9, 4 and 7. As seen in Figure 4.63, gene 2 communicates with six genes 4, 5, 7, 8, 10 and 6. Last, gene 7 interfaces with two genes such as 10, and 8 as clearly seen in the obtained gene network.

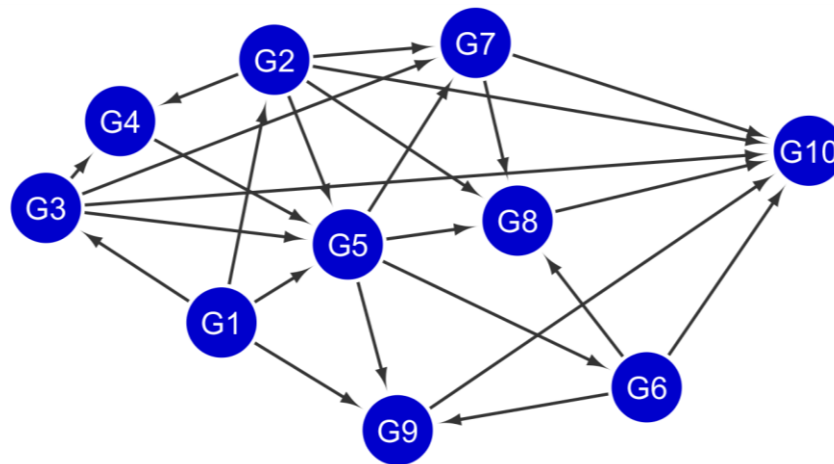


Figure 4.64 Gene network of DREAM4 size 10 network 5 using Pearson method by MRNET.

Source: Cytoscape

This Figure 4.64 is gene network of DREAM4 size 10 network 5 using the Pearson method by the MRNET. In this gene network diagram, there are three gene to gene interactions. For instance, genes 9 and 8 both communicate with gene number 10. Gene 6 connects with three genes, namely 8, 9 and 10. Also, as observed in the network, gene 5 interfaces with four genes 9, 6, 7 and 8. As seen in Figure 4.59, gene 2 communicates with five genes 4, 5, 8, 7 and 10 as clearly seen in the obtained gene network.

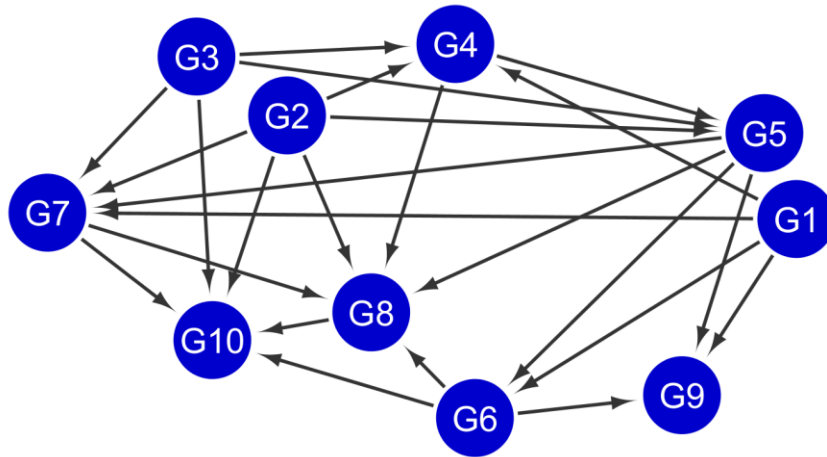


Figure 4.65 Gene network of DREAM4 size 10 network 5 using Spearman method by MRNET.

Source: Cytoscape

This Figure 4.65 is gene network of DREAM4 size 10 network 5 using the Spearman method by the MRNET. In this gene network diagram, there is only one gene to gene interaction. For instance, gene 8 communicates with gene number 10. Gene 4 connects with two genes, namely 5 and 8. Also, as observed in the network, gene 5 interfaces with four genes 8, 9, 7 and 6. Last, as seen in the Figure 4.65, gene 6 interfaces with three genes 8, 9, and 10.

4.3 Comparison of Time-Series Tools Accuracy

The accuracy of the tool is calculated using the following formula:

$$Accuracy = \frac{True\ positives + True\ negatives}{True\ positives + True\ negatives + False\ Negatives + False\ positives} \quad (4.1)$$

In this formula 4.1, components are defined as follow:

- 1) True positives: it means the predicted tool has an edge between genes as well as an edge present in the golden standard.
- 2) True negatives: it means the predicted tool has no edge between genes and no edge present in the golden standard.
- 3) False positives: it means the predicted tool has an edge between genes while no edge is present in the golden standard.
- 4) False negatives: it means the predicted tool has no edge between genes and there is an edge present between genes in the golden standard.

4.3.1 TD-ARACNe

Table 4.1 Accuracy of TD-ARACNe Tool Obtained from DREAM4 Dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	2	1	74	13	0.8444
Network 2	1	3	71	15	0.8
Network 3	3	5	70	12	0.8111
Network 4	4	7	70	9	0.8222
Network 5	4	8	70	8	0.8222

4.3.2 BANJO

Table 4.2 Accuracy of BANJO Tool Obtained from DREAM4 Dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	3	7	68	12	0.7717
Network 2	2	10	64	14	0.7333
Network 3	6	9	67	8	0.8111
Network 4	7	14	63	6	0.7777
Network 5	5	12	66	7	0.7888

4.3.3 Causality

Table 4.3 Accuracy of Causality Tool Obtained from DREAM4 Dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	2	10	64	14	0.7333
Network 2	2	7	67	14	0.7666
Network 3	2	10	65	13	0.7444
Network 4	4	8	69	9	0.8111
Network 5	3	8	70	9	0.8111

The TD-ARACNe performs well on time-series gene expression data while the BANJO and the causality give the same accuracy as seen in the Tables 4.2 and 4.3 respectively

4.4 Comparison of Steady-State Tools Accuracy

4.4.1 ARACNe

Table 4.4 Accuracy of ARACNe Tool using Kendell Algorithm Obtained from DREAM4 dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	3	11	64	12	0.7444
Network 2	4	10	64	12	0.7555
Network 3	5	10	65	10	0.7777
Network 4	1	12	65	12	0.7333
Network 5	3	8	70	9	0.8111

Table 4.5 Accuracy of ARACNe Tool using Pearson Algorithm Obtained from DREAM4 dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	4	6	69	11	0.8111
Network 2	4	8	66	12	0.7777
Network 3	2	10	65	13	0.7444
Network 4	1	12	65	12	0.7333
Network 5	2	7	71	10	0.8111

Table 4.6 Accuracy of ARACNe Tool using Spearman Algorithm Obtained from DREAM4 dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	3	8	67	12	0.7777
Network 2	3	9	65	13	0.7555
Network 3	4	10	65	11	0.7666
Network 4	2	10	67	11	0.7666
Network 5	2	8	70	10	0.8

4.4.2 GENIE3

Table 4.7 Accuracy of GENIE3 Tool Obtained from DREAM4 Dataset

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	7	18	57	8	0.7111
Network 2	3	22	52	13	0.6111
Network 3	6	19	56	9	0.6888
Network 4	2	23	54	11	0.6222
Network 5	6	19	59	6	0.7222

4.4.3 CLR

Table 4.8 Accuracy of CLR Tool using Kendell Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	3	12	64	11	0.7444
Network 2	6	17	57	10	0.7
Network 3	4	12	64	10	0.7555
Network 4	2	12	65	11	0.7444
Network 5	4	17	61	8	0.7222

Table 4.9 Accuracy of CLR Tool using Pearson Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	4	14	61	11	0.7222
Network 2	7	13	61	9	0.7555
Network 3	5	13	62	10	0.7444
Network 4	1	12	65	12	0.7333
Network 5	4	14	64	8	0.7555

Table 4.10 Accuracy of CLR Tool using Spearman Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	3	12	63	12	0.7333
Network 2	6	16	60	10	0.7333
Network 3	4	12	63	11	0.7444
Network 4	2	12	65	11	0.7444
Network 5	4	17	61	8	0.7222

4.4.4 MRNET

Table 4.11 Accuracy of MRNET Tool using Kendell Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	4	16	59	11	0.7
Network 2	7	18	56	9	0.7
Network 3	6	16	59	9	0.7222
Network 4	2	16	61	11	0.7
Network 5	4	17	61	8	0.7222

Table 4.12 Accuracy of MRNET Tool using Pearson Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	4	18	57	11	0.6777
Network 2	7	17	57	9	0.7111
Network 3	6	18	57	9	0.7
Network 4	2	18	59	11	0.6777
Network 5	5	20	58	7	0.7

Table 4.13 Accuracy of MRNET Tool using Spearman Algorithm Obtained from DREAM4

Data	True positives	False positives	True negatives	False negatives	Accuracy
Network 1	5	16	58	11	0.7
Network 2	7	18	56	9	0.7
Network 3	6	17	58	9	0.7111
Network 4	3	19	58	10	0.6777
Network 5	4	21	57	8	0.6777

The ARACNe performs well on steady-state multifactorial data. The CLR is the next tool which performed well after the ARACNe while the GENIE3 is with the least calculated accuracy as seen in the Table 4.7.

CHAPTER 5

CONCLUSION

GRNs play vital role in system biology making it an important area of research. Gene network inference of gene expression data is a challenging task. Hence, many methods have been developed to overcome this problem. The three methods to obtain GRNs are unsupervised, semi-supervised and supervised. This thesis is based on unsupervised methods. GRNs play an important role in system biology which is why more research is performed.

The objective of this thesis was to find the accuracy of different tools and compare their outcomes. This study used two types of data, namely time-series data and steady-state gene expression data to infer gene networks. To obtain GRNs from time-series data, tools such as TD-ARACNe, BANJO, and causality are used while for steady-state data, tools such as ARACNe, GENIE3, CLR and MRNET are used to implicit GRNs. All these tools are based on different algorithms and concepts. As seen in previous sections, TD-ARACNe performs well on time-series gene expression data while BANJO and causality gives the same accuracy. For steady-state data, the ARACNe tool works the best when accuracy was calculated. This gives an idea which tools are more suitable for relevant data available for gene expression.

In retrospect, more commercially available tools could have been used to infer networks to compare the results obtained with greater accuracy.

REFERENCES

- Altay, Gökmen, and Frank Emmert-Streib. "Revealing differences in gene network inference algorithms on the network level by ensemble methods." *Bioinformatics* 26.14 (2010): 1738-1744.
- Bansal, Mukesh, et al. "How to infer gene networks from expression profiles." *Molecular systems biology* 3.1 (2007).
- Bansal, Mukesh, Giusy Della Gatta, and Diego Di Bernardo. "Inference of gene regulatory networks and compound mode of action from time course gene expression profiles." *Bioinformatics* 22.7 (2006): 815-822.
- Basso, Katia, et al. "Reverse engineering of regulatory networks in human B cells." *Nature genetics* 37.4 (2005): 382-390.
- Cerulo, Luigi, Charles Elkan, and Michele Ceccarelli. "Learning gene regulatory networks from only positive and unlabeled data." *Bmc Bioinformatics* 11.1 (2010): 228.
- Faith, Jeremiah J., et al. "Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles." *PLoS biology* 5.1 (2007): e8 Falcon, S., Morgan, M. and Gentleman, R. 2007, An introduction to Bioconductor's ExpressionSet class.
- Faith, Jeremiah J., et al. "Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles." *PLoS biology* 5.1 (2007): e8.
- Filkov, Valdimir. "Identifying gene regulatory networks from gene expression data." *Handbook of Computational Molecular Biology* (2005): 27-1.
- Floratos, Aris, et al. "geWorkbench: an open source platform for integrative genomics." *Bioinformatics* 26.14 (2010): 1779-1780.

- Gardner, Timothy S., Skip Shimer, and James J. Collins. "Inferring microbial genetic networks." *ASM News* 70.3 (2004): 121-126.
- Gautier L. *affydata: Affymetrix Data for Demonstration Purpose*. R package version 1.12.0. (2011)
- Gautier, Laurent, et al. "affy—analysis of Affymetrix GeneChip data at the probe level." *Bioinformatics* 20.3 (2004): 307-315.
- Greenfield, Alex, et al. "DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models." *PloS one* 5.10 (2010): e13397.
- Gregoretti, Francesco, et al. "A parallel implementation of the network identification by multiple regression (NIR) algorithm to reverse-engineer regulatory gene networks." *PloS one* 5.4 (2010): e10179.
- Hauser, Alain, and Peter Bühlmann. "Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs." *The Journal of Machine Learning Research* 13.1 (2012): 2409-2464.
- Hecker, Michael, et al. "Gene regulatory network inference: data integration in dynamic models—a review." *Biosystems* 96.1 (2009): 86-103.
- Irrthum, Alexandre, Louis Wehenkel, and Pierre Geurts. "Inferring regulatory networks from expression data using tree-based methods." *PloS one* 5.9 (2010): e12776.
- Kalisch, Markus, et al. "Causal inference using graphical models with the R package pcalg." *Journal of Statistical Software* 47.11 (2012): 1-26.
- Kalisch, Markus, et al. "Causal inference using graphical models with the R package pcalg." *Journal of Statistical Software* 47.11 (2012): 1-26.
- Karlebach, Guy, and Ron Shamir. "Modelling and analysis of gene regulatory networks." *Nature Reviews Molecular Cell Biology* 9.10 (2008): 770-780.

- Lingeman, Jesse M., and Dennis Shasha. *Network Inference in Molecular Biology: A Hands-on Framework*. Springer Science & Business Media, 2012.
- Madar, Aviv, et al. "DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator." *PloS one* 5.3 (2010): e9803.
- Marbach D, Schaffter T, Mattiussi C. and Floreano D (2009) Generating Realistic *In Silico* Gene Networks for Performance Assessment of Reverse Engineering Methods. *Journal of Computational Biology*, 16(2):229–239.
- Marbach, Daniel, et al. "Generating realistic in silico gene networks for performance assessment of reverse engineering methods." *Journal of computational biology* 16.2 (2009): 229-239.
- Margolin, Adam A., et al. "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context." *BMC bioinformatics* 7.Suppl 1 (2006): S7.
- Margolin, Adam A., et al. "Reverse engineering cellular networks." *Nature protocols* 1.2 (2006): 662-671.
- Meyer, Patrick E., et al. "Information-theoretic inference of large transcriptional regulatory networks." *EURASIP journal on bioinformatics and systems biology* 2007 (2007).
- Meyer, Patrick E., Frederic Lafitte, and Gianluca Bontempi. "minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information." *BMC bioinformatics* 9.1 (2008): 461.
- Peng, Hanchuan, Fulmi Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.8 (2005): 1226-1238.
- S. ZP and M. C (2009). *TDARACNE: Network reverse engineering from time course data*. R package version 1.16.0

- Sagawa, Morihiko, et al. "A new disulfide-linked dimer of a single-chain antibody fragment against human CD47 induces apoptosis in lymphoid malignant cells via the hypoxia inducible factor-1 α pathway." *Cancer science* 102.6 (2011): 1208-1215.
- Saito, Rintaro, et al. "A travel guide to Cytoscape plugins." *Nature methods* 9.11 (2012): 1069-1076.
- Shannon, Paul, et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." *Genome research* 13.11 (2003): 2498-2504.
- Smith, V. Anne, et al. "Computational inference of neural information flow networks." *PLoS computational biology* 2.11 (2006): e161.
- Stolovitzky, Gustavo, Don Monroe, and Andrea Califano. "Dialogue on reverse-engineering assessment and methods-The DREAM of high-throughput pathway inference." *REVERSE ENGINEERING BIOLOGICAL NETWORKS* 1115 (2007): 1-22.
- Stolovitzky, Gustavo, Robert J. Prill, and Andrea Califano. "Lessons from the DREAM2 Challenges." *Annals of the New York Academy of Sciences* 1158.1 (2009): 159-195.
- Verbanck, Marie, Sébastien Lê, and Jérôme Pagès. "A new unsupervised gene clustering algorithm based on the integration of biological knowledge into expression data." *BMC bioinformatics* 14.1 (2013): 42.
- Zavlanos, Michael M., et al. "Inferring stable genetic networks from steady-state data." *Automatica* 47.6 (2011): 1113-1122.
- Zoppoli, Pietro, Sandro Morganella, and Michele Ceccarelli. "TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach." *Bmc Bioinformatics* 11.1 (2010): 154.