Evaluation of network inference algorithms and their effects on network analysis for the study of small metabolomic data sets

by

Haley Greenyer
B.Sc., University of Vermont, USA, 2019

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Evaluation of network inference algorithms and their effects on network analysis for the study of small metabolomic data sets

by

Haley Greenyer
B.Sc., University of Vermont, USA, 2019

Supervisory Committee

Hosna Jabbari, Supervisor
Department of Computer Science

Dr. Ulrike Stege, Supervisor
Department of Computer Science

ABSTRACT

Motivation: Alzheimer's Disease (AD) is a highly prevalent, neurodegenerative disease which causes gradual cognitive decline. As documented in the literature, evidence has recently mounted for the role of metabolic dysfunction in AD. Metabolomic data has therefore been increasingly used in AD studies. Metabolomic disease studies often suffer from small sample sizes and inflated false discovery rates. It is therefore of great importance to identify algorithms best suited for the inference of metabolic networks from small cohort disease studies. For future benchmarking, and for the development of new metabolic network inference methods, it is similarly important to identify appropriate performance measures for small sample sizes.

Results: The performances of 13 different network inference algorithms, including correlation-based, regression-based, information theoretic, and hybrid methods, were assessed through benchmarking and structural network analyses. Benchmarking was performed on simulated data with known structures across six sample sizes using three different summative performance measures: area under the Receiver Operating Characteristic Curve, area under the Precision Recall Curve, and Matthews Correlation Coefficient. Structural analyses (commonly applied in disease studies), including betweenness, closeness, and eigenvector centrality were applied to simulated data. Differential network analysis was additionally applied to experimental AD data. Based on the performance measure benchmarking and network analysis results, I identified Probabilistic Context Likelihood Relatedness of Correlation with Biweight Midcorrelation ($PCLRC_b$) (a novel variation of the $PCLRC$ algorithm) to be best suited for the prediction of metabolic networks from small-cohort disease studies. Additionally, I identified Matthews Correlation Coefficient as the best measure with which to evaluate the performance of metabolic network inference methods across small sample sizes.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

My co-supervisors **Hosna Jabbari** and **Ulrike Stege** for encouraging me to follow my passions and supporting me throughout this process.

The members of the **Computational Biology Research and Analytics** group at the University of Victoria for their continual encouragement and valuable feedback.

DEDICATION

I would like to dedicate my thesis to Andrew Hollar, without whom I would never have found my way to this work.

Chapter 1

Introduction

1.1 Alzheimer's Disease

Alzheimer's Disease (AD) is a slowly progressive neurodegenerative disease that typically presents as a loss of episodic memory and gradual cognitive decline [85]. AD is the most common cause of dementia and is best characterized by neuronal loss and accumulation of β-amyloid ($A\beta$) and twisted strands of tau protein in the brain [33]. Tau protein is the major component of neurofibrillary tangles [17, 116] while $A\beta$ has been shown to be the major component of senile plaques [32, 76]. It has been speculated that these plaques and tangles impair neural-network activity and damage the synapses between neurons that form and maintain microcircuits that support a variety of cognitive functions [85]. For an in-depth review of the evidence implicating $A\beta$ and tau protein in neurodegeneration, see [33].

AD can be broadly classified as either sporadic AD (sAD) or early-onset AD (EOAD). EOAD is defined as beginning before age 65 and comprises about 5-6% of all AD cases [82]. EOAD is more aggressive than sAD and arises from a mutation in the amyloid precursor protein (APP), or the PSEN-1 or PSEN-2 genes. Each of these mutations results in increased production of $A\beta$ [16, 64]. No explicit genetic link has been established as the cause of sAD, though individuals carrying the $\epsilon 4$ allele of apolipoprotein E have increased risk of developing AD at a younger age [66]. Current studies predominantly support the Amyloid hypothesis, which proposes that sAD is determined by genetic instability and is initiated by the accumulation of $A\beta$ [96].

For both sAD and EOAD, age is the greatest risk factor [30]. It is believed that

neurodegeneration begins up to 20 years before clinical onset [85]. Even in the case of transgenic mice specifically engineered to develop EOAD, ageing is required to observe pathology [38]. This indicates that some mechanism of ageing is required for neurodegeneration and that the expression of tau protein and $A\beta$ alone are not sufficient cause for disease [1].

Evidence has recently been mounting for the importance of metabolic decline in AD pathology [11, 10, 59], with some studies suggesting that AD should be classified as a metabolic disease [54, 22, 19]. Metabolism is the set of chemical reactions in which *metabolites* (small molecules) act as substrates or products. These enzyme-catalyzed reactions allow organisms to grow, reproduce, and respond to their environments. Metabolism is known to alter with age, and such alterations are now being considered potential causes of AD [54, 5]. The ageing process can be characterized metabolically by insulin resistance, changes in body composition, and physiological decline of growth hormones, insulin-like growth factor-1 (IGF-1), and sex steroids [5]. Through review of the role of metabolism in AD, Kang et al. suggested that metabolites such as insulin and reactive oxygen species could be therapeutic targets for developing effective AD treatments [54]. For a more in-depth review of the role of metabolism in ageing, AD, and other neurodegenerative diseases, see [54, 11, 107, 5].

1.2 Metabolomics and Lipidomics

The ability to accurately measure multiple metabolites directly from complex biological systems has brought forth metabolomics as a promising lens through which to study AD and other human diseases [27]. The metabolome can be defined as the set of all small molecule metabolites found in a specific cell, organ, or organism [117]. Metabolomics detects changes in the metabolome, which represents an accurate biochemical profile of the organism in healthy and diseased states [108]. Metabolite profiling, which quantifies the levels of metabolites in a sample, is an application of metabolomics that has routinely been used on biological fluids (biofluids) as a way to characterize the health state of a patient [98]. The availability of biorepositories, which contain biofluids and tissue samples from control and AD patients, provide an outstanding opportunity to advance the understanding of sex- and disease-specific metabolic signatures and mechanisms [115]. Additionally, with blood as a readily available biofluid for recurrent measures, longitudinal studies using metabolomics could significantly enhance the precision of individualized medicine [115].

Lipidomics is a sub-discipline of metabolomics that focuses on the study of cellular lipids (water-insoluble metabolites), which comprise the lipidome. Lipids are highly dynamic small molecules that play many essential roles in cellular function, including cellular barriers, membrane matrices, signaling, and energy depots [121, 41]. Given the strong association between lipids and metabolic dysfunction-related diseases, lipidomics can be an important tool for mechanistic studies, risk prediction, and therapeutic monitoring [79].

Metabolic and lipidomic pathways are largely conserved between species. Mice share 99% of genes with humans, and mouse models have therefore become a useful tool with which to study AD and other human diseases [90]. Though EOAD accounts for a small number of cases, mouse models that have been genetically altered to express APP, PS1, and PS2 transgenes have advanced the ability to study AD mechanisms [115, 56]. Though no existing mouse model perfectly recreates the complexity of AD as observed in humans, they have substantially aided the study of AD and AD drug discovery [3]. Together, metabolomic and lipidomic profiling has provided strong evidence that multiple factors are involved in the pathology of AD [115].

Meaningful results can be extracted from metabolic and lipidomic data through the use of network theory [51, 63]. However, previous work has suggested that animal studies of neurological disorders are often underpowered as a result of small sample size and a large number of features, among other factors [26, 97, 110]. Underpowered studies are at greater risk for false positive findings. It is therefore of great importance to evaluate the performance of network inference algorithms for the specific application of predicting networks from small sample size metabolic data. If no existing algorithms are determined to perform well, it would indicate the need for the development of a new algorithm for this specific application.

1.3 Thesis Objective

The overall objective of this thesis is to contribute to existing literature by identifying the methods best suited for the inference of metabolic networks from small sample sizes. In order to identify such algorithms, I first review existing network inference methods commonly applied to biological studies (see Chapter 2). Of these reviewed methods, I select those which have either previously been shown to outperform other algorithms of a similar type, or have not yet been benchmarked on

metabolic data for inclusion in performance analyses (see Section 3.2.1). Three main analyses are conducted to investigate network inference algorithm performance. The first consists of benchmarking all selected algorithms using three different summative performance measures against a simulated data set with known structures (see Section 3.1.1). The second consists of comparing the structural characteristics of predicted networks against a simulated reference network. The third analysis consists of comparing the results of Differential Connectivity Analysis (see Section 2.2) when applied to Alzheimer's disease networks inferred using different algorithms. Based on results from benchmarking and the two structural analyses, I identify the algorithm best suited for the inference of metabolic networks from small-cohort disease studies ($PCLRC_b$, as defined in Section 2.1.5). I also identify the performance measure best suited for the evaluation of metabolic network inference algorithms across small sample sizes (MCC, as defined in Section 2.1.1). Identification of this performance measure will improve the validity of future benchmarking efforts and aid in the development of new algorithms specifically for the inference of metabolic networks.

1.4 Thesis Contributions

Through completion of the this thesis, I made the following contributions:

1. I simulated a metabolic data set from a dynamic metabolism model [48].

2. Using this simulated data set with known structures, I benchmarked the performance of 13 network inference algorithms across six sample sizes for three different summative performance measures.

3. Using both the simulated data and experimental Alzheimer's data [35, 36], I examined the effects of selected algorithms on three different node centrality measures, as well as on Differential Connectivity Analysis (see Section 2.2 for definitions).

4. I present two novel variations, of the highly regarded Probabilistic Context Likelihood of Relatedness of Correlation algorithm, $PCLRC_k$ and $PCLRC_b$ (see Section 2.1.5 for the definition of $PCLRC$), for the inference of metabolic networks. Both variations were found to outperform the original implementation based on the aforementioned analyses.

5. Based on benchmarking and structural analysis results, I identified $PCLRC_b$ as the algorithm best suited overall for the prediction of metabolic networks from small sample sizes for disease studies.

6. Through examination of predicted networks and their structural properties, I identified Matthews Correlation Coefficient (see Section 2.1.1 for definition) to be the performance measure best suited for the assessment of metabolic network inference methods across small sample sizes.

7. I identify the need for a new metabolic network inference algorithm that preserves structural properties.

8. I identify the need for a new summative performance measure that considered the structural characteristics of predicted networks.

Chapter 2

Background

In this chapter, I provide background information for a wide variety of network inference algorithms and network analyses. Section 2.1.1 begins with a review of the measures commonly used to assess the performance of network inference methods. The terms defined in this section will be referenced throughout the chapter in reference to previous studies. In following Sections 2.1.2 – 2.1.5, I review network inference methods that are well-regarded for the inference of biological networks. The reviewed algorithms are grouped into four method types: correlation-based, information theoretic, regression-based, and hybrid. The chapter concludes with Section 2.2, in which I review several analyses commonly used to investigate the properties of biological networks in the study of human diseases.

2.1 Biological Networks

In recent decades, network science has been increasingly recognized as a fundamental tool for the investigation of complex biological systems [89]. The inference and subsequent analysis of biological networks has contributed to discoveries in several fields including neuroscience [29], proteomics [109, 75], transcriptomics [71], and metabolomics [102, 34]. Biological networks are abstractions of molecular entities and their interactions, which are represented as nodes and edges, respectively. Edges are established based on some measure of similarity between biological entities.

In metabolic networks, nodes represent individual metabolites measured in biological samples and edges are established between metabolites based on some measure of similarity. The similarity between two metabolites x and y is calculated from a paired collection of concentrations. A given metabolite x can be represented as $x = [x_1, x_2, ..., x_n]$, where instances $1, 2, ...n$ represent different biological samples from which the concentration of x was measured. Inference of metabolic association networks requires paired data of the form: (x_1, y_1),...,(x_n, y_n), where x_n and y_n are the concentrations of metabolites x and y, respectively, measured from biological sample n. This data format is assumed for all method descriptions unless otherwise specified. In its simplest form, the adjacency matrix A of a weighted metabolic association network can be defined as follows:

$$A_{x,y} = S(x, y) \tag{2.1}$$

where for each entry $A_{x,y}$ of A, x and y represent a given pair of metabolites, and S represents a similarity measure produced by a given network inference algorithm. When some threshold τ is applied to the entries of A, a binary (un-weighted) network B can be obtained:

$$B_{x,y} = \begin{cases} 1 \text{ if } S(x, y) > \tau \\ 0 \text{ otherwise} \end{cases} \tag{2.2}$$

Metabolic association patterns are known to change upon perturbation of the system, which could be induced experimentally or result from the onset of pathological conditions [106, 62, 100]. As such, reconstruction of metabolite interactions through network inference can facilitate investigation of the interactions between different metabolites in response to disease. The majority of algorithms currently used to infer

metabolic association networks were originally developed for the prediction of gene regulatory networks. However, since the focus of this thesis is metabolomics, the descriptions of all algorithms included in the following subsections will be formulated with respect to the inference of metabolic networks (i.e., x and y will denote metabolites rather than genes). For a more general review of biological network inference methods, see [48, 89].

2.1.1 Performance Measures

A variety of measures exist in the literature for the evaluation and comparison of network inference methods. Within this chapter I discuss previous studies which compare the performance of various network inference algorithms. I will therefore provide context and definitions for all referenced measures in the following section.

It should be noted that several of algorithms discussed were previously assessed in network inference challenges from the Dialogue for Reverse Engineering Assessments and Methods (DREAM) project [103]. Specifically, the fourth and fifth set of DREAM systems biology challenges, DREAM4 and DREAM5, which both asked participants to predict gene regulatory networks from a variety of data sources. Both of these challenges rely on some combination of the following measures to evaluate algorithm performance.

Confusion Matrix

When a reference network containing 'true' interactions is available, binary networks (see Equation 2.1 above) can be assessed as binary classifications. Four results can occur when comparing predicted classifications to a reference: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These results can be represented as a confusion matrix (see Figure 2.1).

A variety of measures can be derived from the confusion matrix for the assessment of binary classifications. The following subsections describe a number of these measures that have been selected for later benchmarking. Each subsection provides a description of the given measure, as well as justification for its inclusion in this study. All uses of P, N, PP, PN, TP, FP, TN, and FN in subsequent formulas and definitions correspond to their meanings within the Confusion Matrix shown in Figure 2.1.

		Predicted Condition	
	Total *= P + N*	Positives (PP) *= 1 classification*	Negatives (PN) *= 0 classification*
Reference Condition	Positives (P) *= 1 classification*	True Positive (TP)	False Negative (FN)
	Negatives (N) *= 0 classification*	False Positive (FP)	True Negative (TN)

Figure 2.1: Confusion Matrix illustrating all possible results of binary classification. A *1 classification* refers to a 1 entry in the binary network adjacency matrix (present edge) and *0 classification* indicates a 0 entry in the binary network adjacency matrix (absent edge). TP is a correctly predicted positive condition, FP is a falsely predicted positive condition, TN is a correctly predicted negative condition, and FN is a falsely predicted negative condition.

False Discovery Rate and Positive Predictive Value

One issue plaguing existing network inference methods is the inability to distinguish between direct and indirect relationships, resulting in a large number of false positive classifications. It is therefore desirable to minimize the number of false positives produced by these methods. The false discovery rate (FDR) can be defined as the likelihood that a predicted positive will be false. FDR is calculated as:

$$FDR = \frac{FP}{PP} = \frac{FP}{FP + TP} \tag{2.3}$$

The positive predictive value (PPV) is the complement of FDR, and can be defined as the likelihood that predictive positives will be true. PPV can be calculated as:

$$PPV = \frac{TP}{PP} = \frac{TP}{TP + FP} \tag{2.4}$$

Receiver Operating Characteristic Curve

The ROC curve is a plot depicting the trade-off between *True Positive Rate* (TPR) and *False Positive Rate* (FPR), explicit formulas for which are provided below:

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{2.5}$$

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} \tag{2.6}$$

TPR defines how many correct positive classifications are captured in the predicted network, while the FPR defines how many positive classifications are falsely captured in the predicted network. The best achievable value for TPR is 1, while the best achievable value for FPR is 0.

To construct the ROC curve for a given network inference method, TPR and FPR (Equations 2.5 and 2.6) are calculated for a range of thresholds applied to a weighted adjacency matrix. The lower limit of this range is the minimum value present in the (weighted) network, while the upper limit is the maximum value present in the network. Once calculations have been performed for all thresholds in the range, paired TPR and FPR observations can be plotted against one another (with TPR along the y axis and FPR along the x axis, as shown in Figure 2.2).

Area under the ROC curve (AUROC) is a widely used measure of classification performance that has often been applied to diagnostic tests [74, 120, 8, 61, 39]. It has been previously shown that while AUROC can discriminate well between 'bad' and 'good' models, it does not discriminate well between different 'good' models [74]. AUROC has also been shown to perform poorly on imbalanced data (i.e., sparse binary networks, which are common in metabolomics) [7, 123, 40].

Precision-Recall Curve

The precision-recall (PR) curve is a popular alternative to the ROC curve for the assessment of binary classifiers [18]. The PR curve is a plot depicting the trade off between PPV (also known as precision, see Equation 2.4) and TPR (also known as recall, see Equation 2.5). The PR curve can be constructed in the same manner as a ROC curve, with PPV and TPR calculated in place of TPR and FPR, respectively. Paired PPV and TPR observations can then be plotted against each other as shown in Figure 2.3.

Similar to AUROC, area under the PR curve (AUPR) can be calculated to obtain a summative statistic for the performance of a binary classifier. Due to the imbalanced nature of binary biological networks, AUPR is becoming increasingly favored over AUROC as a network inference performance measure in bioinformatics [94].

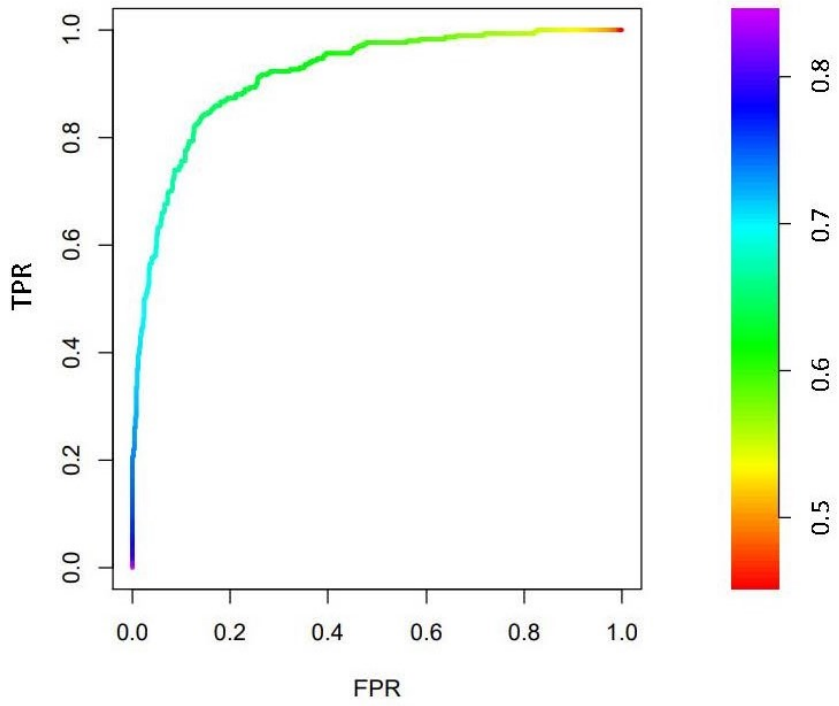

Figure 2.2: Example of a ROC curve. A range of thresholds, indicated by the color bar, are applied to a weighted network. For each binary network produced, TPR and FPR are calculated against a reference binary network. The paired TPR and FPR values are then plotted against one another. Area under the ROC curve can be used as a summative measure of performance. This figure was produced using the **R** package *PRROC* [37].

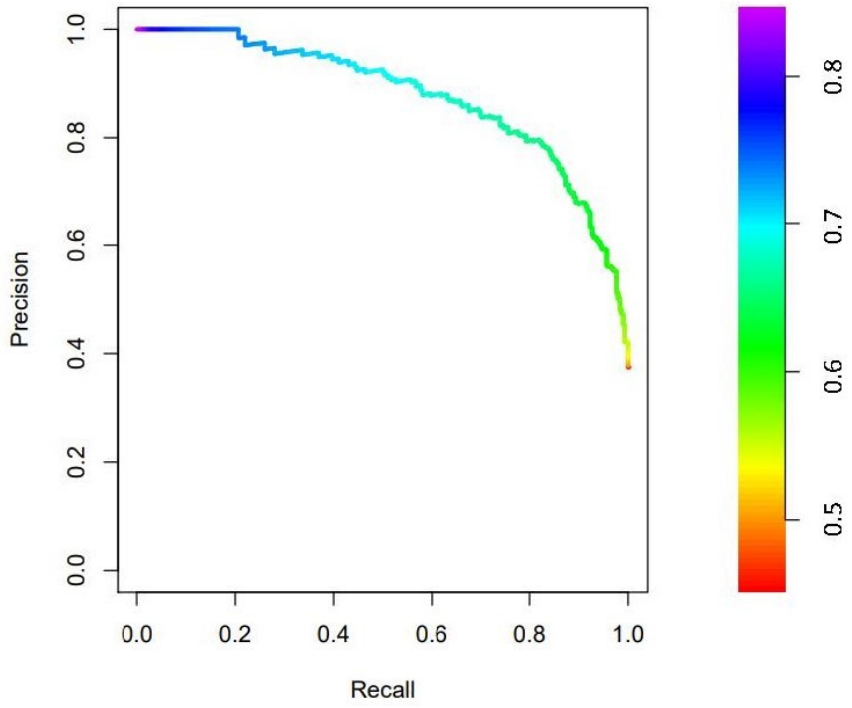

Figure 2.3: Example of a PR curve. A range of thresholds, indicated by the color bar, are applied to a weighted network. For each binary network produced, precision (PPV) and recall (TPR) are calculated against a reference binary network. The paired precision and recall values are then plotted against one another. Area under the ROC curve can be used as a summative measure of performance. This figure was produced using the **R** package *PRROC* [37].

Matthews Correlation Coefficient

Matthews Correlation Coefficient (MCC) [77] is a widely used performance metric in binary classification[13, 7, 14]. MCC has been shown to outperform AUROC in the assessment of network inference methods on imbalanced data [7, 123, 40]. MCC measures the correlation between the reference and predicted conditions (i.e., P and N, 1 and 0) [14, 13]:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2.7}$$

MCC can range in value from [-1,1], with -1 indicating poor performance and +1 indicating strong performance. An MCC of 0 indicates that the algorithm performs as a random classifier.

2.1.2 Correlation-based Methods

Correlation is one of the simplest and most commonly used measures for quantifying the level of association between a pair of metabolites. Correlation-based methods are generally computationally inexpensive and therefore scalable for large data sets [93]. Weighted metabolic association networks can be obtained by applying a correlation measure ($CORR$) to each possible pairing of metabolites in a data set, resulting in an adjacency matrix of correlation coefficients. Binary networks can be obtained by applying a threshold τ to the correlation coefficients $CORR(x, y)$ and/or a threshold α to the corresponding p-values to select edges, resulting in the binary adjacency matrix A wherein:

$$A_{x,y} = \begin{cases} 1 \text{ if } |CORR(x, y)| > \tau \text{ and/or } p_{x,y} < \alpha, \\ 0 \text{ otherwise} \end{cases} \tag{2.8}$$

A variety of correlations have been applied to the inference of biological networks; predominantly Pearson's and Spearman's correlations ($CORR_p$ and $CORR_s$, respectively). These, along with Kendall's correlation ($CORR_k$), Biweight Midcorrelation ($bicor$) and partial correlation ($pcor$) are discussed in the following subsections. Camacho et al. indicated that when used in metabolomic studies, correlations should be determined with large samples (e.g., larger than 10 times the number of variables present in the data) using $CORR_s$ [12]. When fewer samples are available, it is recommended that $CORR_p$ be used with a high significance threshold [12]. In a

study conducted by Santos et al., the strengths of several correlation methods were evaluated on simulated and experimental gene expression data. Their results suggest that if linearity or monotonicity can be assumed, $CORR_s$ or $CORR_k$ may be more useful than $CORR_p$, because of their ability to identify both linear and non-linear monotonic relationships with high power [21].

Pearson's Correlation

Pearson's correlation ($CORR_p$) assumes a normal data distribution and can be used to identify linear relationships [86]. Given paired data, Pearson's correlation coefficient $CORR_p(x, y)$ can be calculated for a given pair of metabolites x and y as:

$$CORR_p(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.9}$$

where n is the sample size, x_i and y_i are individual paired samples, and $\bar{x}$ and $\bar{y}$ are the sample means of x and y, respectively. When perfect linear dependence is observed, $CORR_p(x, y) = \pm 1$. When complete linear independence is observed, $CORR_p(x, y) = 0$.

Spearman's Correlation

Spearman's correlation ($CORR_s$) is a non-parametric test used to measure the degree of association between two variables [101]. $CORR_s$ can be considered an extension of $CORR_p$ wherein the data is converted to ranks before calculating the coefficient. Spearman's correlation makes no assumptions of the underlying data distribution and can detect monotonic relationships. The Spearman's correlation coefficient $CORR_s(x, y)$ for a given pair of metabolites x and y can be calculated as:

$$CORR_s(x,y) = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{2.10}$$

Where n is the sample size and d_i is the distance between the ranks of individual paired samples x_i and y_i. Similar to Pearson's correlation coefficient, a value of $CORR_s(x, y) = \pm 1$ indicates a perfect monotonic relationship between x and y while $CORR_s(x, y) = 0$ indicates monotonic independence.

Kendall's Correlation

Similar to Spearman's correlation, Kendall's correlation ($CORR_k$) is a ranked correlation test that can be used to identify monotonic relationships [55]. Kendall's correlation is less sensitive to error in the data than Spearman's and has better p-values with smaller sample sizes. For a given pair of metabolites x and y, Kendall's correlation coefficient $CORR_k(x, y)$ can be calculated as:

$$CORR_k(x,y) = \frac{\left\{(\#\ concordant\ pairs) - (\#\ discordant\ pairs)\right\}}{0.5n(n-1)} \tag{2.11}$$

For a given data pair (x_i, y_i) and (x_j, y_j), the pair is *concordant* if either both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. The pair is *discordant* if either $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$.

Biweight Midcorrelation

Biweight Midcorrelation (*bicor*) is considered to be an alternative to $CORR_p$ that is more robust against outliers [114]. *bicor* has previously been successfully applied to the inference of gene regulatory networks and the investigation of differential coexpression [122]. The coefficient $bicor(x, y)$ for metabolites x and y can be calculated as follows [114]:

Two vectors $u = [u_1, u_2, ...u_n]$ and $v = [v_1, v_2, ...v_n]$ are first defined where for each vector element $i = 1, ..., n$:

$$u_i = \frac{x_i - med\,(x)}{9 \cdot mad\,(x)} \tag{2.12}$$

$$v_i = \frac{y_i - med\,(y)}{9 \cdot mad\,(y)} \tag{2.13}$$

where $med(*)$ is the median and $mad(*)$ is the absolute deviation of the metabolite concentrations. This leads to definition of the weights $w_i^{(x)}$ for x_i and $w_i^{(y)}$ for y_i:

$$w_i^{(x)} = \left(1 - u_i^2\right)^2 \mathbf{I}\,(1 - |u_i|) \tag{2.14}$$

$$w_i^{(y)} = \left(1 - v_i^2\right)^2 \mathbf{I}\,(1 - |v_i|) \tag{2.15}$$

where **I** is an indicator function that takes on the following values:

$$\mathbf{I}(1-|u_i|) = \begin{cases} 1, \ if \ (1-|u_i|) > 0 \\ 0 \text{ otherwise} \end{cases} \tag{2.16}$$

$$\mathbf{I}(1-|v_i|) = \begin{cases} 1, \ if \ (1-|v_i|) > 0 \\ 0 \text{ otherwise} \end{cases} \tag{2.17}$$

Therefore, the value of weights $w_i^{(x)}$ and $w_i^{(y)}$ range from 0 to 1. Based on these weights, the Biweight Midcorrelation of metabolites x and y can be calculated as:

$$bicor(x,y) = \frac{\sum_{i=1}^{m}(x_i - med(x))w_i^{(x)}(y_i - med(y))w_i^{(y)}}{\sqrt{\sum_{j=1}^{m}[(x_j - med(x))w_j^{(x)}]^2}\sqrt{\sum_{k=1}^{m}[(x_k - med(x))w_k^{(x)}]^2}} \tag{2.18}$$

Partial Correlation

Generally, correlations observed between metabolites result from a combination of all reactions and regulatory processes in the network. Metabolic networks contain fast biochemical reactions in an open mass-flow system, giving way to indirect, system-wide correlations between distantly connected metabolites [58]. One limitation of correlation-based network inference is the inability to distinguish between direct and indirect interactions. Several studies have investigated the use of partial correlation (*pcor*) to mitigate these spurious associations. Partial correlation measures the correlation between two variables after their linear dependence upon other variables is removed, thus allowing distinction between direct and indirect associations [2]. Kumsiek et al. used Gaussian graphical models (GGMs), which can be interpreted as partial correlation coefficients, to infer a metabolic network from a large human population cohort and found that high *pcor* coefficients generally correspond to known metabolic reactions [58]. However, partial correlations are far more computationally expensive and thus may be impractical as a base measure of association for methods that incorporate resampling.

2.1.3 Information Theoretic Methods

The framework of information theory, first developed by Claude Shannon in 1948, provides a general measure of dependency between variables. Information theoretic methods most often use mutual information (MI), an information theory concept, as a base measure of similarity for network inference. Similar to CORR methods, MI can be used on its own to produce weighted networks, with thresholding applied to produce binary networks. The following section further describes MI and several information theoretic network inference algorithms, which combine MI and other information theory concepts, including Context Likelihood of Relatedness (CLR), The Multicast Reduction NETwork ($MRNET$) algorithm, and the Algorithm for the Reconstruction of Accurate Cellular Networks ($ARACNE$).

Mutual Information

Mutual Information (MI) provides a measure of mutual dependence between two random variables. The mutual information $MI(x, y)$ between two metabolites x and y can be calculated as follows [60]:

$$MI(x, y) = \sum_{i,j}^{n} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{2.19}$$

where $p(x_i, y_j)$ is the joint probability distribution function of x_i and y_j and $p(x_i)$ and $p(y_j)$ are the probabilities that $x = x_i$ and $y = y_i$, respectively. Entries of a weighted adjacency matrix generated using MI can have values in the range of $[0, +\infty]$.

MI presents the advantage of capturing non-linear relationships, which are known to be present in biological systems [49]. A wide range of MI-based methods have been developed for the inference of biological networks, including $ARACNE$ [73], $MRNET$ [83], and CLR [28]. For additional discussion and evaluation of MI-based methods see [48, 91].

CLR

The context likelihood of relatedness (CLR) algorithm uses MI as a base similarity measure between biological entities. The CLR algorithm estimates the likelihood of the MI score for a given pair of metabolites by testing that score against a background MI distribution, thus minimizing the effects of indirect interactions [28]. For a given

pair of metabolites x and y, $MI(x, y)$ is compared to a joint background distribution constructed from the set of all MI scores for x and the set of all MI scores for y (MI_y). The final form of the likelihood estimate is [28]:

$$CLR(x, y) = f(z_x, z_y) = \sqrt{z_x^2 + z_y^2} \tag{2.20}$$

where

$$z_i = max\left\{0, \frac{MI(i, k) - \mu_i}{\sigma_i}\right\} \tag{2.21}$$

and μ_i and σ_i are the mean and the standard deviation of the empirical distribution of $MI(ik)$, with $k = 1, ..., n$, where n is the total number of metabolites. CLR results in a weighted adjacency matrix wherein each entry gives a measure of probability of association between variables. Zero entries indicate that $MI(x, y)$ was not found to be greater than the background distribution mean.

The CLR algorithm was originally developed for the inference of gene regulatory networks using MI, but has since been adapted to accept other measures of similarity as input. In an effort to increase robustness against levels of error and noise that often accompany metabolic data and small sample sizes, Saccenti et al. extended the CLR algorithm by replacing MI with correlation and introducing resampling procedures [92] (see Section 2.1.5).

MRNET

The Multicast Reduction NETwork ($MRNET$) algorithm is an information theoretic method originally developed for the inference of gene regulatory networks from microarray data [83]. $MRNET$ evaluates an MI adjacency matrix based on the maximum relevance/minimum redundancy (MRMR) algorithm [24], a feature selection technique that has been successfully applied to disease biomarker identification [52, 70]. The MRMR algorithm iteratively selects the feature that has maximum relevance with respect to the target variable and minimum redundancy relative to the features that have been selected in previous iterations [24, 87].

Each metabolite in the set of metabolites m is treated as the target variable m_x in turn, and all other metabolites are then treated as candidate features. The MI score for m_x and the candidate metabolite m_i are calculated, and the candidate m_i with the largest MI with the target m_x is added to the set S. In each subsequent

step, candidate metabolites m_i are selected and added to S through maximization of Equation 2.22:

$$s_i = u_i - r_i \tag{2.22}$$

where

$$u_i = MI(m_i, m_x) \tag{2.23}$$

$$r_i = \frac{1}{|S|} \sum_{m_k \in S} MI(m_i, m_k) \tag{2.24}$$

In Equations 2.22–2.24 above, u_i represents the relevance between m_i and m_x, and r_i measures the redundancy between m_x and previously selected candidates. For each pair of metabolites m_i and m_k, two scores s_i and s_j are returned. $MRNET$ selects the maximum of these two scores to score the interaction between the two metabolites. In its original publication, $MRNET$ was shown to perform competitively against CLR and $ARACNE$ on 30 simulated microarray data sets [83]. Song et al. later included $MRNET$ in a benchmarking study comparing the performance of several MI-based network inference algorithms on eight empirical data sets [99]. $MRNET$ was found to perform comparably to $ARACNE$ and CLR, outperforming both in only one application.

ARACNE

The Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) is a network inference method originally developed to infer regulatory networks in mammalian cells from gene expression data [73]. Given a weighted MI adjacency matrix, $ARACNE$ excludes indirect interactions by considering edge triplets in the network. For each triplet, the edge with the weakest weight (lowest MI score) is pruned (set to 0 in the adjacency matrix) if the difference between the two lowest scores is above a cutoff value ϵ. For a given triplet of edges xy, xz, and yz between nodes x, y, and z, the edge xy is pruned if:

$$MI(x, y) \leq \min\left(MI(x, z), MI(y, z)\right) \tag{2.25}$$

The performance of $ARACNE$ has previously been assessed on simulated data for

what were defined to be a large sample size of 500 and a small sample size of 50 [104]. Benchmarked against CLR, CORR, and $PCLRC$, it was shown that $ARACNE$ is not suitable for the inference of metabolite-metabolite association networks from small sample sizes [104].

2.1.4 Regression-based Methods

Regression-based methods seek to quantify the relationship between a given pair of metabolites by solving a regression problem in order to predict one from the other. This method presents the advantage of predicting causal relationships, but may fail to detect non-linear interactions due to the assumptions of linearity made by some algorithms [93]. Two notably popular regression-based algorithms, $GENIE3$ [45] and $TIGRESS$ [43], are described in this section.

GENIE3

$GENIE3$ (GEne Network Inference with Ensemble of trees) is a network inference algorithm based on variable selection with ensembles of regression trees. Because $GENIE3$ uses tree-ensemble-based methods, it has the potential to detect non-linear relationships. The $GENIE3$ framework can be outlined in terms of metabolic network prediction as follows [45]:

A multifactoral learning sample from which to infer a network as a sample of n measurements can be defined as:

$$LS = \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$$

where $\mathbf{x}_k \in \mathbb{R}^p$ for $k = 1, ..., n$ is a vector of concentration values of all p metabolites in the k^{th} biological sample:

$$\mathbf{x}_k = (x_k^1, x_k^2, ..., x_k^p)^T$$

The vector $\mathbf{x}_k^{-j}$ containing the concentration values in the kth biological sample for all metabolites except metabolite j can be written as:

$$\mathbf{x}_k^{-j} = (x_k^1, ..., x_k^{j-1}, x_k^{j+1}, ..., x_k^p)^T$$

It then follows that:

$$x_k^j = f_j(x_k^{-j}) + \varepsilon_k, \forall k$$

where ε_k is a random noise with zero mean, conditionally to $\mathbf{x}_k^{-j}$. The assumption is made that the function f_j only exploits the concentration in $\mathbf{x}^{-j}$ of the metabolites that directly regulate metabolite j.

1. For each metabolite $j = 1, ..., p$, a learning sample LS^j is generated, where the learning sample of input-output pairs for metabolite j is denoted:

$$LS^j = (\mathbf{x}_k^{-j}, x_k^j), k = 1, ...N$$

2. A function f_j is learned from LS^j and a local ranking of all genes except j is computed. The function f_j minimizes the following error:

$$\sum_{k=1}^{N} (x_k^j - f_j(\mathbf{x}_k^{-j}))^2$$

   which is solved using regression by developing tree structured models.

3. The p local rankings are aggregated to get a global ranking of all regulatory links.

$GENIE3$ was developed for the inference of directed gene regulatory networks, and was the overall top performer based on AUPR in the DREAM4 *In Silico Multifactoral* challenge for the reconstruction of the *Escherichia coli* network [45, 72].

TIGRESS

The Trustful Inference of Gene REgulation using Stability Selection (TIGRESS) algorithm employs a similar framework as $GENIE3$, formulating network inference as a collection of feature selection problems for each target variable [43]. The main difference between $TIGRESS$ and $GENIE3$ is the aggregation of feature selection. Where $GENIE3$ uses decision tress, $TIGRESS$ uses a combination of least angle regression (LARS) [25] (a popular multivariate feature selection method) and stability selection [43]. Stability selection consists of repeatedly running LARS, resampling the samples and variables at each run, and computing the frequency with which each variable is selected across all runs [80]. $TIGRESS$ was among the top performing gene regulatory network prediction algorithms included in the DREAM5 challenge, and was determined to be the best overall linear regression-based method [72]. In its original publication, the performance of $TIGRESS$ was evaluated via AUPR and AUROC against several other network inference algorithms, including CLR, $ARACNE$,

and $GENIE3$, on nine gene expression benchmark data sets (consisting of both simulated and experimental data). In this study, $TIGRESS$ did not achieve overall better performance than $GENIE3$ [45].

2.1.5 Hybrid Algorithms

Many network inference methods have been proposed that combine several different approaches. Based on results of the DREAM5 challenge, in which over 30 network inference algorithms were assessed on *Escherichia coli, Staphylococcus aureus*, and *in silico* microarray data, Marbach et al. presented a community-based method for gene regulatory network inference [72]. Observing that different methods have complementary advantages and limitations under different contexts, they integrated the predictions of all assessed algorithms. Though hybrid methods performed more robustly across data sets than other categories of methods, they did not match the robustness and performance of the community predictions [72]. For the purpose of this thesis, only stand-alone algorithms are of interest. However, it should be noted that the community-based predictor was the overall best performing method in the DREAM5 challenge [72].

The DREAM5 challenge, though specific to gene regulatory networks, made a strong case for the utility of hybrid predictors for the inference of biological networks. One hybrid network inference algorithm specifically designed for the inference of metabolite-metabolite association networks, $PCLRC$ [92], has found growing popularity in recent years. The details of this algorithm and its previous applications are discussed in the following section.

PCLRC

The Probabilistic Context Likelihood of Relatedness of Correlation (PCLRC) algorithm combines techniques from correlation-based and information theoretic methods. The $PCLRC$ algorithm is an extension of CLR that was developed for the inference of metabolic association networks [92, 28]. It extends CLR by substituting $MI(x, y)$ with $CORR_p(x, y)$ and by introducing resampling. MI, $CORR_p$, and $CORR_s$ have all been assessed as the similarity measure in $PCLRC$, with multiple studies indicating Pearson's correlation performs best [92, 48].

Pearson's correlation is used in the original presentation of the algorithm [92]. In each iteration of $PCLRC$, 75% of the data samples are randomly selected from

which to generate a $CORR_p$ matrix. CLR is then performed on the correlation matrix to generate a probabilistic measure of likelihood for each correlation coefficient. A dynamic threshold is then applied to the CLR matrix that retains the top $Q\%$ most likely interactions (pruning all others). This process is repeated for k iterations, with the resulting edge weights indicating the proportion of iterations for which each edge was retained. To obtain a binary network, a final threshold is imposed to include only those edges that were retained in at least 95% of iterations. Pseudo-code for this algorithm is shown below using default values for Q and k from the original implementation [92].

Algorithm 1 PCLRC

Input: *$n \times m$ data matrix with n samples and m variables*

$edgeCounts \leftarrow m \times m$ *matrix*
$k \leftarrow 10^5$
$Q \leftarrow 0.30$
$sampleSize \leftarrow 0.75$

for i=1 to k **do**
  $sampleMatrix \leftarrow randomSample(data\ matrix, sampleSize)$
  $corrMatrix \leftarrow CORR(sampleMatrix)$
  $clrMatrix \leftarrow CLR(corrMatrix)$
  $thresh \leftarrow getThreshold(clrMatrix, Q)$ {get threshold value for the top Q%}
  $clrMatrix[\ clrMatrix < thresh\ ] \leftarrow 0$
  $clrMatrix[\ clrMatrix \geq thresh\ ] \leftarrow 1$
  $edgeCounts = edgeCounts + clrMatrix$
end for

Return: $edgeCounts \div k$

$PCLRC$ has been used in several metabolomic studies. When presenting the algorithm, Saccenti et al. used $PCLRC$ to investigate latent cardiovascular risk in a cohort of relatively healthy and young blood donors [92]. Vignoli et al. later used $PCLRC$ to successfully identify metabolic determinants associated with mortality in Acute Myocardial Infraction patients by predicting networks associated with different patient risk factors [111]. The introduction of iterative sampling in $PCLRC$ minimizes the number of spurious edges and increases robustness of the predicted interactions.

2.2 Network Analysis

A biological network can be defined as a *graph* $G = (V, E)$ where V is the set of nodes (i.e., biological entities) and E is the set of edges (i.e., interactions between entities). Once constructed, biological networks can be analyzed using algorithms from graph theory. Metabolic association networks can be represented as *undirected* graphs wherein an edge between metabolites i and j is defined as $E = \{(i, j)|\ i, j \in V\}$. For a weighted network, this set of edges E is associated with a weight function $w : E \longrightarrow \mathbb{R}$ where the weight w_{ij} of an edge between nodes i and j indicates the strength of association. The measures and algorithms for network analysis discussed in this section include only those directly relevant to the subject matter of this thesis (i.e., can be applied to undirected metabolic networks or gene regulatory networks).

2.2.1 Network Structure

In network analysis, node centrality can be used to answer questions such as: which node is the most 'important', which node behaves as a hub, or which node acts as a bridge between different communities [57]. Different measures of centrality have been used to address such questions.

The degree of a node v_i, also known as *degree centrality*, can be defined as the number of edges connected to v_i. Many biological networks, including metabolic networks, are known to exhibit a scale-free characteristic where node degrees follow a power-law distribution. This distribution results in a structure containing many weakly connected nodes and few highly connected nodes. Scale-free networks are known to be very robust to perturbations [4].

One centrality measure that has often been applied to biological networks is *betweenness*. The betweenness centrality quantifies the influence a given node has over the flow of communication between other nodes. Nodes with a high betweenness and a low degree have been known to support network modularity in other contexts such as protein interaction networks [67]. The betweenness of a given node v_i is defined as [31]:

$$BC(i) = \sum_{i,j,k \in V, j \neq k} \frac{\delta_{j,k}(i)}{\delta_{j,k}} \tag{2.26}$$

where $\delta_{j,k}$ is the number of the shortest paths from node j to node k and $\delta_{j,k}(i)$ is the number of these shortest paths that pass through node v_i.

Another measure of centrality, known as *closeness*, is used to detect important nodes which can communicate quickly with other nodes in a network. The closeness centrality of a given node v_i can be calculated as the mean value of the inverse of the distance from v_i to other nodes [67]:

$$CC(i) = \frac{1}{N-1} \sum_{i,j \in V, j \neq i} \frac{1}{d_{ij}} \tag{2.27}$$

where N is the number of nodes in the graph, and the distance between two nodes d_{ij} is defined as the shortest path (minimum number edges to be traversed) between them. If no path exists between i and j, $d_{ij} = 0$. Nodes with larger closeness centralities will be more central within the network and therefore should exchange information more efficiently with other nodes. Closeness has previously been used to identify important metabolites in the glycolysis and citrate acid cycle pathways in *E.coli* [69].

Both betweenness and closeness have been applied to metabolic networks to investigate changes in network structure that arise from environmental changes as well as the evolution of network organization across different species [78].

Where degree-based centralities are only based on the number of a node's neighbors, eigenvector-based centralities measure the influence of a node by considering the influence of a node's neighbors. The eigenvector centrality of a node v_i can be calculated as the sum of it's neighbors' centralities [6]:

$$EC(i) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{i,j} EC(j) \tag{2.28}$$

where $A_{i,j}$ is an entry in the network adjacency matrix A and λ is the largest eigenvalue of A. Eigenvector centrality has been successfully applied in a variety of areas including as a tool for mapping the brain network in various states [68]. All measures of centrality suffer from different sensitivities and a single centrality measure is not sufficient to accurately identify essential nodes in biological networks.

2.2.2 Differential Network Analysis

Multiple networks can be constructed to represent the same biological systems under various conditions (i.e., disease and control). Differential network analysis explores differences in the structural characteristics of two or more networks. This analysis

can be applied to reveal affected molecular mechanisms between different conditions [50, 46]. The general procedure for *differential network analysis* can be outlined as follows:

1. Construct two networks G_1 and G_2, from data sets D_1 and D_2, that contain the same set of metabolites and represent two distinct network states to (e.g., control and disease).

2. Compute the desired structural measure Θ (e.g., degree, betweenness, etc.) for all nodes in G_1 and G_2, resulting in two vectors, Θ_1 and Θ_2, of length n, where n is the total number of nodes.

3. Compute the difference between the measures calculated for each network: $D_0 = |\Theta_1 - \Theta_2|$. The resulting vector D_0 is of length n.

4. Conduct a permutation test to assess the statistical significance of each entry in D_0.

where the procedure for a permutation test with P permutations can be outlined as follows:

Initialize a count vector C of length n. Then for P permutations:

1. Obtain two new permuted data sets D_{1p} and D_{2p} by resampling the rows of D_1 and D_2 without replacement.

2. Construct networks G_{1p} and G_{2p} from D_{1p} and D_{2p}, respectively.

3. Compute Θ_{1p} and Θ_{2p} from G_{1p} and G_{2p}, respectively.

4. Compute the difference between the measures calculated for the nodes of each permuted network: $D_p = |\Theta_{1p} - \Theta_{2p}|$.

5. For each node n_i, if $D_p[n_i] > D_0[n_i]$, add 1 to $C[n_i]$.

where p-values corresponding to the observed differences between networks can be obtained by dividing each entry of the count vector C by the number of permutations performed, P. These p-values indicate the proportion of permuted differences that were found to be larger than the actual difference. A smaller p-value indicates that the observed difference is not likely to have occurred by random chance.

Changes in centrality measures between networks may indicate changes in the network 'importance' of specific biological entities was altered by the change in condition. Differential connectivity analysis (DCA) is a variation of differential network analysis commonly applied to metabolic networks that tests for differences in node degree between states [47, 111, 91]. For example, Vignoli et al. used DCA to successfully identify metabolic determinants associated with mortality in Acute Myocardial Infarction patients [111]. DCA has been shown to be a useful tool for the analysis of biological networks in human disease studies.

Chapter 3

Materials and Methods

In this chapter I describe the various data sets and measures used to evaluate the performance of selected network inference algorithms. The chapter begins with the description of a data set simulated from a dynamic model in Section 3.1, upon which all benchmarking was performed. Also described in Section 3.1 is the experimental mouse model Alzheimer's data set used to further explore the structure of predicted networks. Section 3.2 follows with a detailed overview of the benchmarking procedure used to compare the performance of selected algorithms. Details regarding the comparison of structural analyses are also provided in this section. The chapter concludes with a list of all software and packages used to simulate data samples, process experimental data, infer networks, and calculate performance measures.

3.1 Materials

This section details the characteristics of two distinct data sets. The first data set was simulated from a dynamic metabolism model following a procedure presented by Jahagirdar et al. [48]. Section 3.1.1 provides a biological overview of the dynamic model used to simulate these data samples, followed by a description of the data simulation process. This simulated data was used for all benchmarking calculations. The second data set, described in Section 3.1.2, contains lipid panels obtained from a mouse-model Alzheimer's experiment presented by Granger et al. [35, 36] . This experimental data set is used for the inference and comparison of networks representing different disease states.

3.1.1 Simulated Data

Previous studies have highlighted the utility of simulated data with known structures for the evaluation of network inference algorithm performance [48, 92, 72]. Based on the results of their study on the reconstruction of metabolic association networks, Jahagirdar et al. advocated for the use of dynamic models to generate simulated data over standard multivariate approaches [48]. In order to compare the performance of several algorithms for the inference of small sample size metabolic networks, data samples were simulated from the Arachadonich Acid (AA) dynamic model presented in [48]. A detailed account of construction of the dynamic model, optimization of parameters, and simulation of metabolite concentration profiles are provided in the aforementioned study. For the sake of completeness, those details relevant to this thesis are recounted below.

Dynamic Model of Arachidonic Acid Metabolism

AA and its derivatives (eicosanoids) have been extensively studied in recent decades due to their activity as signalling molecules and their role in a vast number of inflammatory conditions [9, 113, 42]. Jahagirdar et al. developed a dynamic model of AA degradation [48] by retrieving information about metabolites and eiconasoids (lipids) involved in AA metabolism, as well as associated reactions from the following databases: Recon2.2 [105], The Human Metabolome Database (HMDB) [117], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [53] and the Chemical Entities of Biological Interest database (ChEBI) [20]. Kinetic data, such as reaction and equilibrium constants, was collected from BRENDA [95] and SABIO-RK [118]. The constructed AA metabolic model includes 83 metabolites and 131 reactions. An ordinary differential equation was written for each metabolite in the model, with enzymatic reactions described using Michaelis-Menten kinetics [84, 48]:

$$v = \frac{V_{max} \times [S]}{K_m + [S]} \tag{3.1}$$

where v is the reaction rate, V_{max} is the maximum reaction rate achieved by the system, [S] is the substrate concentration, and K_m is the Michaelis-Menten constant. Non-enzymatic reactions are described by simple mass action laws [48]:

$$v = k_i \times [S] \tag{3.2}$$

where k_i, a proportionality constant, is the equilibrium constant of the reaction.

A reference network was obtained using the stoichiometric matrix of the AA metabolism model, which indicates what reaction each metabolite participates in. This reference network is represented as a binary adjacency matrix of primary reactions, where entries of value 1 indicate direct interaction and entries of value 0 indicate no interaction or indirect interaction. The resulting reference network $\mathbf{N}_{ref}$ is shown in Figure 3.1.

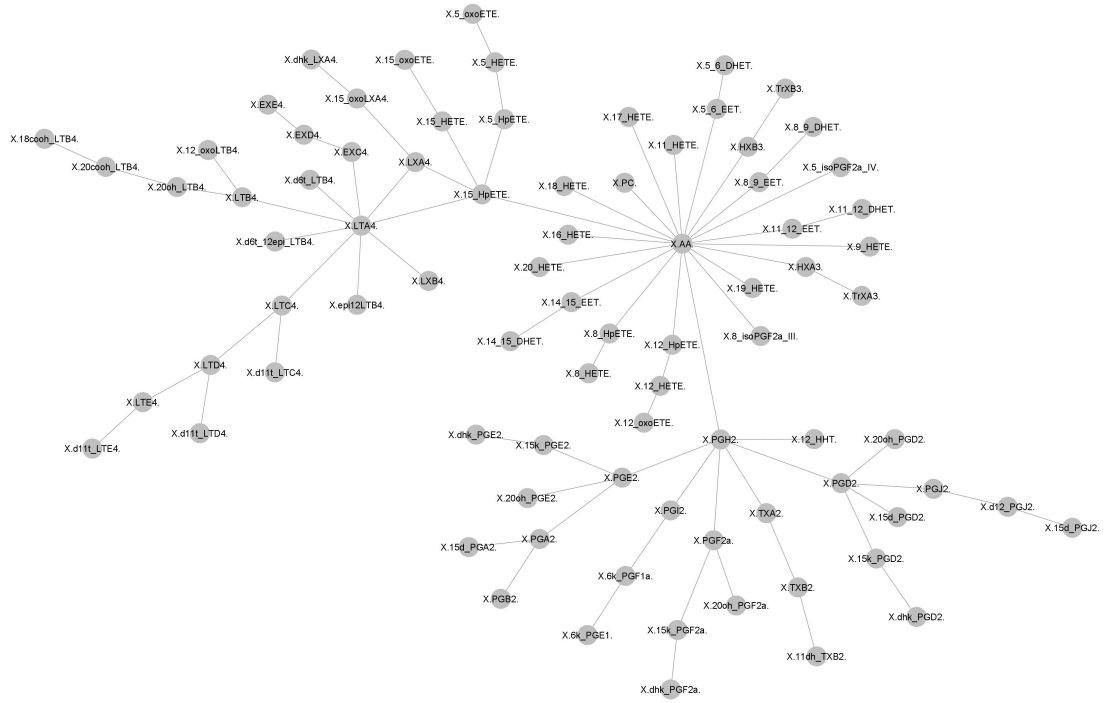

Figure 3.1: Reference network ($\mathbf{N}_{ref}$) representing primary interactions of the dynamic model of AA metabolism. Nodes represent metabolites. Each node is labeled with the corresponding metabolite Connections between nodes indicate a direct reactionary interaction.

Simulation of Metabolite Concentration Profiles

Data samples were simulated from the dynamic AA model using the COmplex PAthway SImulator (COPASI). COPASI is a singular application for simulation and analysis of biochemical pathways and their dynamics [81, 44]. COPASI has been widely used in research modeling of biological systems and in the development of theory and computational methods (See `https://copasi.org/Research`).

Using COPASI 4.34 (build 251), $n = 500$ data samples (rows) were simulated from the AA model, where each sample contains a concentration (nmol/L) for each of the 83 metabolites (columns). Each individual sample was produced by running a single time course simulation on the AA model, during which metabolite concentrations were sampled at 90h after induction of AA conversion. This time point within the simulation was selected to achieve a steady state condition. Experimental results have previously reported that a period of 70h after induction of AA conversion is required to achieve steady state [23, 88]. Variation between profiles was achieved by slightly perturbing the model parameters (constants defined in Equations 3.1 and 3.2). Before running each time course, values for these parameters were randomly sampled from a uniform distribution with limits of +/- 10% of the parameters' optimized values. An explicit procedure for the generation of n data samples from the AA model in COPASI can be outlined as follows:

To simulate *n* data samples in COPASI:

1. Configure the *Time Course* Task such that output occurs at 90h.
2. Configure the *Parameter Scan* Task to repeat n times.
3. In the *Parameter Scan* Task, add new scan item: *Time Course*.
4. In the *Parameter Scan* Task, add new scan item: *Random Distribution* for all Reactions present in the model, and select *uniform* as the distribution type.
5. Set the lower and upper *Random Distribution* limits to $0.9 \times$ *optimized value* and $1.1 \times$ *optimized value*, respectively, for each reaction parameter.
6. Execute the *Parameter Scan Task*

3.1.2 Experimental Data

In addition to the simulated data, the effects of different network inference methods were compared on experimental data obtained from a mouse model Alzheimer's study [35, 36]. This experimental data was included to further demonstrate the characteristic differences between networks predicted using different inference methods.

TgCRND8 Mouse Model

The amyloid cascade (see Section 1.1) has been modeled in transgenic mice expressing the human familial mutant ($A\beta PP$) and/or PS1 genes linked to EOAD. The TgCRND8 (Tg) line is an aggressive EOAD mouse model that encodes a double mutant form of $A\beta PP$ 695 under the control of the Prp gene promoter [15]. This model has previously been used to study the sexual dimorphisms of AD [35] and AD-linked changes in glycerophosphocholine metabolism [36].

The TgCRND8 data set used in this study contains lipid panels (235 lipids) extracted from sibling male and female mice. Out of 20 male mice, 7 have a Tg genotype. Out of 20 female mice, 10 have a Tg genotype, for a total of 17 Tg mice and 23 non-Tg mice. To obtain the lipid panels, blood was drawn from the animals between 343 and 366 days of age. Samples were collected from each animal in fasted and non-fasted states, for a total of 80 samples from 40 animals.

The great genetic similarity between animals in this study allows for greater confidence in results obtained from small sample sizes. Though the conclusions drawn from mouse model studies do not directly translate into conclusions in human pathology, a much larger sample size of human data would be needed to account for confounding factors such as genetic diversity, medications, or diet.

Data Pre-processing

Lipids found in levels too small to quantify in the TgCRND8 data set resulted in missing (NA) entries. Lipids found to be NA in more than 60% of samples were removed, leaving 199 lipids in the data set. Any remaining NA entries were then replaced with five-fold less than the minimum value recorded for the corresponding lipids.

Following the removal and replacement of NA entries, Z-score standardization and log-transformation were applied across all entries. This achieves lipid-wise zero-mean and unit-variance, which ensures all lipids are equally weighted in downstream analyses and corrects skew, respectively. An explicit procedure for this data pre-processing can be written as follows:

1. Remove lipids with NA entries in $> 60\%$ of samples.
2. Replace remaining NA entries with $0.2 \times$ (observed minimum lipid level).
3. Apply log-transformation.
4. Apply Z-score standardization .

For the purpose of this study, two subsets were extracted from the full data set for network prediction:

1. Fasted, Tg mice (fTg): Mice with the Tg genotype whose lipid panels were collected while in a fasted state
2. Fasted, Non-Tg mice ($fnTg$): Mice with a non-Tg (control) genotype whose lipid panels were collected while in a fasted state

These subsets were selected to facilitate investigation of the effects of different network inference algorithms on DCA (see Section 2.2.2).

3.2 Methods

The following section details all analyses used to evaluate the performance of selected network inference algorithms on simulated and experimental data. This includes benchmarking conducted using three different summative performance measures on simulated, and structural analyses performed on experimental data.

3.2.1 Benchmarking

Selected network inference algorithms were benchmarked using three different measures across six different sample sizes:

$$n = [5, 10, 20, 50, 100, 200]$$

Network inference algorithms were included in benchmarking based on their performance against other algorithms in previous studies, as well as potential suitability for small sample size network inference (as discussed in Section 2.1). Methods previously demonstrated to perform poorly on small sample sizes were excluded from consideration.

The algorithms included in benchmarking can be grouped as follows. From correlation-based methods: Pearson's Correlation $CORR_p$, Spearman's Correlation $CORR_s$, Kendall's Correlation $CORR_k$, Biweight Midcorrelation (*bicor*), and partial correlation estimated from GGMs (*pcor*). All correlation-based methods were included for benchmarking because $CORR_p$ and $CORR_s$ remain popular and widely used methods for the inference of biological networks, and to the best of my knowledge, the other aforementioned methods have not previously been evaluated against one another for this application.

From information theoretic methods: MI, CLR, and $MRNET$ were included in benchmarking. $ARACNE$ was not included as it has previously been shown to perform poorly on small sample sizes (see Section 2.1.3).

From regression-based methods, only $GENIE3$ was included. $TIGRESS$ was excluded from benchmarking because $GENIE3$ has been shown to outperform $TIGRESS$ in multiple studies according to multiple performance measures (see Section 2.1.4).

Lastly, for hybrid methods, four variations of PCLRC were included, each variation using a different base correlation measure: Pearson's ($PCLRC_p$), Spearman's ($PCLRC_s$), Kendall's ($PCLRC_k$), and Biweight Midcorrelation ($PCLRC_b$). To the best of my knowledge, variations of $PCLRC$ using *bicor* and $CORR_k$ have not previously been presented or assessed for this application. A *pcor* variation of PCLRC was not considered in this study due to the combined computational complexity of both algorithms. In total, 13 methods were included in benchmarking.

The performance of each selected network inference algorithm was evaluated for sample sizes: $n = [5, 10, 20, 50, 100, 200]$. For the purpose of this study, 'small' sample sizes refer to sample sizes smaller than the number of variables present in the data set. In reference to the simulated data set (which includes 83 variables), small sample sizes are considered to be: 5, 10, 20, and 50. Similarly, 'large' sample sizes are those greater than or equal to the number of variables present in the data set. Varied sample sizes were generated by randomly sampling rows (samples) from the full 500-sample simulated data set.

For each algorithm, AUORC, AUPR, MCC, FDR, and PPV were calculated using predicted networks and N_{ref} (see Section 3.1.1) for each sample size. Of these measures, AUC and AUPR do not require thresholding. To calculate measures requiring a threshold to produce binary adjacency matrices, an 'optimal' threshold was determined based on the corresponding ROC curve.

The 'optimal' threshold (τ) was considered to be the threshold that resulted in maximized TPR and minimized FPR on each algorithm's ROC curve. In other words, the selected threshold optimized the proportion of the AA reference network correctly classified in the predicted networks. A threshold was extracted specific to each algorithm and applied to obtain binary networks from which MCC, FDR, and PPV could be calculated. All calculations were repeated and averaged over 100 runs. The overall benchmarking procedure can be outlined as follows:

1. For each sample size s:
   (a) Obtain a data set D_s of sample size s by randomly sampling s rows from simulated data set D.
   (b) For each network inference method M:
      i. Construct a ROC curve for M using D_s and network N_{ref}, calculate the AUROC, and determine an optimal threshold τ.
      ii. Construct PR curve for M using D_s and N_{ref}, calculate the AUPR.
      iii. Infer weighted network N_W from D_s using M. Apply τ to N_W obtain binary network N_B.
      iv. Calculate MCC, FDR, and PPV from N_B and N_{ref}
2. Repeat step 1 for 100 iterations and average the results.

Based on within-group scores, the 'best' performing algorithm was defined as the algorithm achieving the highest average performance measure score across the majority of small sample sizes: 5, 10, 20, and 50 (regardless of calculated statistical significance). Should no algorithm be distinguished by this criteria, the same criteria was extended to large sample sizes of 100 and 200. The top performing algorithms from each group (along with GENIE3, as the only regression-based method) were then directly compared. Of these top performers, the algorithm that showed the most statistically significant improvement over other methods across small sample sizes was determined to be the overall top performing algorithm.

3.2.2 Permutation Test

To determine whether the differences in performance measure (Θ) observed between benchmarking results for a given pair of algorithms $A1$ and $A2$, a permutation test was conducted as outlined below:

1. Calculate the difference between the performance measures Θ_{A1} and Θ_{A2} for algorithms $A1$ and $A2$: $\delta_{ref} = |\Theta_{A1} - \Theta_{A2}|$

2. For $n = 1000$ iterations:

   (a) Randomly rearrange the rows (samples) for each column (metabolite) of the data set D, with which benchmarking was conducted, producing permuted data set D_P.

   (b) Infer networks N_{A1} and N_{A2} using algorithms $A1$ and $A2$, respectively, from D_P

   (c) For permuted networks N_{A1} and N_{A2}, calculate measures Θ_{PA1} and Θ_{PA2}, and the difference between them: $\delta_P = |\Theta_{PA1} - \Theta_{PA2}|$

3. calculate the proportion p of n iterations for which $\delta_p > \delta_{ref}$

The proportion p returned from the end of the permutation test acts as a p-value upon which a threshold can be placed to determine significance. For this study, a threshold of $p < 0.05$ was used to determine statistically significant difference between calculated metrics for different network inference methods.

3.2.3 Structural Analysis

To further investigate the effects of top performing algorithms on network reconstruction, structural analyses were applied to networks inferred from both simulated and experimental data. To illustrate the effects of these algorithms at small sample sizes, a small subset (sample size $n = 20$) was randomly sampled from the full simulated data set. Networks were predicted from this subset using the top two performing algorithms (as determined through benchmarking), resulting in two weighted networks: N_{A1} and N_{A2}. Three different node centralities, betweenness, closeness, and eigenvector centrality (described in Section 2.2.1) were calculated for N_{A1}, N_{A2}, and the AA metabolism reference network N_{ref}. Visualizations of each network were produced for each different centrality measure. Within these network visualizations, nodes were colored based on their magnitude of centrality score, allowing for direct visual comparisons between the node centralities of N_{A1}, N_{A2}, and N_{ref}. For each network, the 20 nodes with the highest of each centrality measure were also extracted fora more explicit direct comparison, since large centrality scores have previously been associated with biological significance (see Section 2.2.1)

A comparative structural analysis, DCA, was performed on experimental data. For each of the subsets extracted from the TgCRND8 data set (see Section 3.1.2) a weighted network was predicted using each of the top performing algorithms A1 and A2, resulting in four networks (two fTg and two $fnTg$). Following the procedure described in Section 2.2.2, DCA was performed on each pair of fTg and $nfTg$ networks.

3.3 Software

COPASI 4.34 (build 251) was used to simulate all test case metabolic profiles, following the procedure described in Section 3.1.1. **R** version 4.1.0 (see `https://www.r-project.org`) was used to perform all network inference, performance metric calculations, and network analyses. Correlation algorithms $CORR_p$, $CORR_s$, and $CORR_k$ were performed using the **R** package *stats* (version 3.6.2). The bicor algorithm was performed using the *bicor()* function from the **R** package *WGCNA* (version 1.70-3). The *pcor* method was performed using the *ggm.estimate.pcor()* function from the **R** package *GeneNet* (version 1.2.16). All PCLRC algorithm variations were performed using a modified version of an **R** implementation from [91]. This implementation was modified to allow for substitution between base correlation measures used in the algorithm ($CORR_p$, $CORR_s$, $CORR_k$, and bicor). MI was calculated using *mutualInfoAdjacency()* from the **R** package *WGCNA*. Information theoretic algorithms CLR and MRNET were performed using the functions *clr()* and *mrnet()* from the **R** package *minet* (version 3.30.0). Finally, regression-based method GENIE3 was performed using the **R** implementation from *Bioconductor* (version 3.14). Benchmarking was performed by calculating all measures for each algorithm in parallel using an 8-node *Azure Batch* pool with *Standard_D1_v2* sized virtual machines.

Chapter 4

Results

In this chapter I present my findings on the evaluation of network inference algorithm performance for the inference of small sample size metabolic networks. In Section 4.1 I compare the differences in performance presented by different assessment measures AUROC, AUPR, and MCC (for definitions, see Section 2.1.1). I primarily focus on the implications of the observed results for small sample sizes (defined in Section 3.2), which are common in mouse-model Alzheimer's studies and clinical studies. Section 4.1 concludes with a discussion of the suitability of AUROC, AUPR, and MCC as summative performance measures for this application. Topological effects of the two best overall performing algorithms (identified in Section 4.1) are further explored in Section 4.2. Networks were predicted for structural analysis from both experimental and simulated data. For all networks predicted from simulated data, the structural properties of betweenness, closeness, and eigenvector-centrality (defined in Section 2.2.1) were calculated. The results of DCA (see Section 2.2.2), performed on networks predicted from the experimental data, are discussed in Section 4.2.2. This chapter concludes with a final comparison of the networks predicted using different network inference algorithms and recommendations for which algorithms and measures should be used for this application.

4.1 Network Inference Algorithm Performance

As described in Section 3.2, the performances of 13 different network inference algorithms were assessed on simulated metabolic data (for details, see Section 3.1.1). The algorithms selected for assessment can be broken into four groups: correlation-

based methods, information theoretic methods, regression-based methods, and hybrid methods. The benchmarking results for AUROC, AUPR, and MCC are summarized in Figures 4.1, 4.2, and 4.3, respectively. In each of these figures, Subfigure **a** visualizes the results for correlation-based methods, Subfigure **b** visualizes the results for *PCLRC* variations, Subfigure **c** visualizes the results for information theoretic methods, and Subfigure **d** compares the results of top performers across groups. See Section 3.2.1 for details regarding the identification of top performing algorithms. Within these figures, labels placed above bars that do not share letters indicates statistically significant difference between scores, as determined by permutation test (for details, see Section 3.2.2). An asterisk placed above a label indicates that algorithm has performed statistically significantly better than all other algorithms in that group.

4.1.1 AUROC

With respect to AUROC, *bicor* achieved the highest average score for the majority of sample sizes, statistically significantly outperforming all other correlation-based algorithms for the small sample size 50, and large sample sizes of 100 and 200 (shown in Figure 4.1**a**). To the best of my knowledge, *bicor* has not been benchmarked against other correlation-based methods (or any other type of method) for this application. $CORR_p$ and $CORR_s$ methods have previously been benchmarked for sample sizes of 50 and 500 on simulated data generated from the same AA metabolism dynamic model [48]. The AUROC results presented here for $CORR_p$ and $CORR_s$ for the sample size of 50 agree with the results of this previous study.

Of the *PCLRC* variations, $PCLRC_b$ achieved the highest average AUROC for the majority of sample sizes, statistically significantly outperforming all other variations for the small sample size of 50, and large sample sizes of 100 and 200 (see Figure 4.1**b**). Although $PCLRC_b$ achieved the highest average AUROC score for small sample size 20, no statistically significant difference was found between $PCLRC_b$, $PCLRC_k$, and $PCLRC_s$ scores. Notably, the original *PCLRC* algorithm, denoted $PCLRC_p$, achieved the lowest average AUROC across almost all sample sizes.

Amongst the information theoretic methods, CLR achieved the highest AUROC across all sample sizes other than 5 (see Figure 4.1**c**). However, no statistically significant difference was observed between CLR, MI, and MRNET at any sample size other than 20, where CLR and MI are both shown to statistically significantly outperform MRNET.

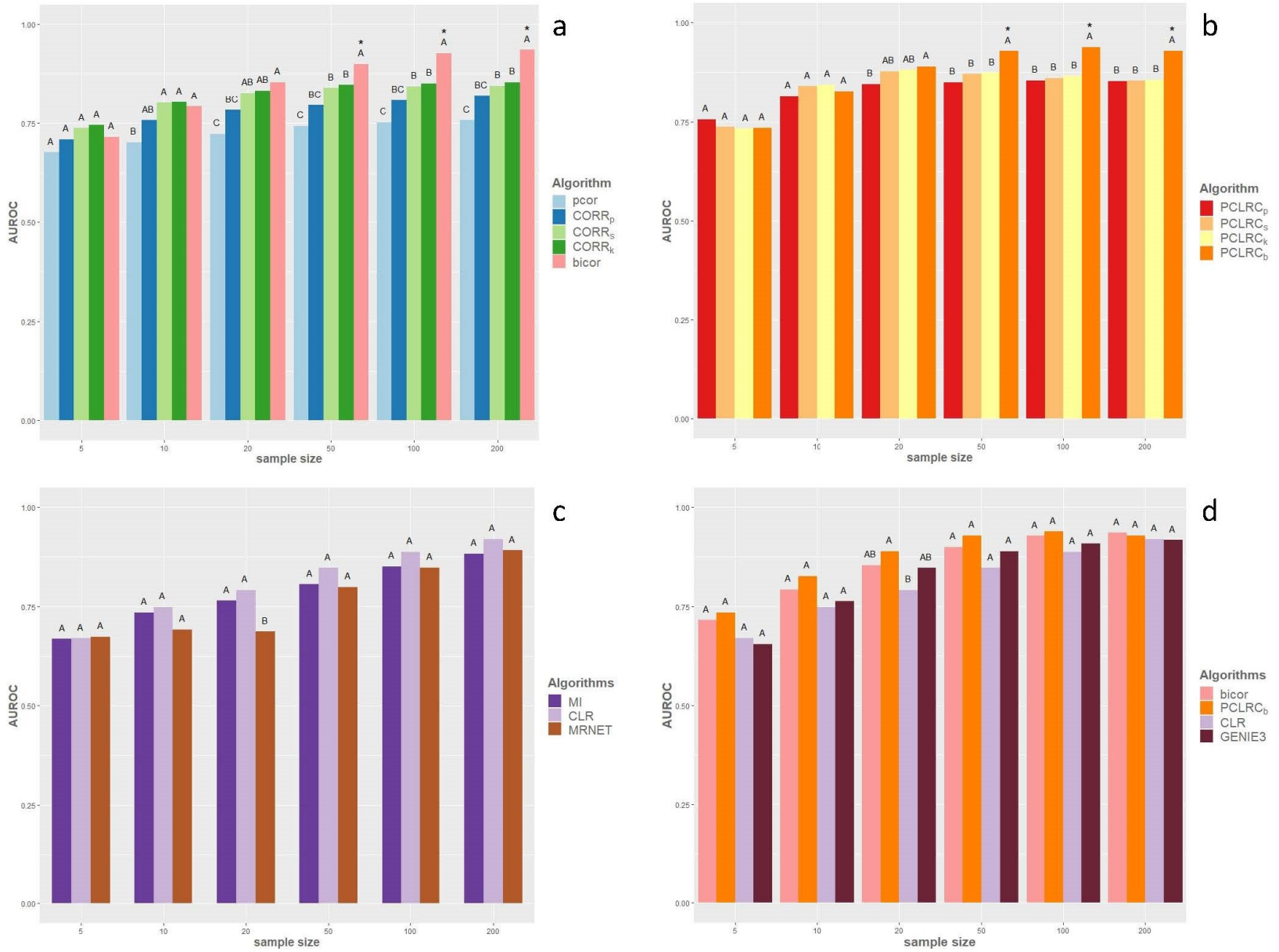

Figure 4.1: AUROC values averaged over 100 runs for various network inference methods. Performance was evaluated against simulated data with known structures for a range of sample sizes. Different letters denote a statistically significant difference ($p < 0.05$) among averages at each sample size according to a two-sided permutation test. Asterisks indicate methods which performed statistically significantly better than all other methods for a given sample size. Subfigure **a** compares AUROC values for correlation-based methods. Subfigure **b** compares AUROC values for *PCLRC* variations. Subfigure **c** compares AUROC values for Information Theoretic methods. Subfigure **d** compares AUROC values for the top performing methods from subfigures **a**-**c**

Directly comparing the top performing algorithms from each group (see Figure 4.1**d**), all methods performed comparably for the small sample size of 50, and large sample sizes of 100 and 200. Only $PCLRC_b$ showed statistically significant improvement over any method, performing better than CLR at sample size 20. $PCLRC_b$ also achieved the highest average AUROC score across sample sizes of 10, 20, 50 and 100, indicating $PCLRC_b$ as the top performing method across small sample sizes. It should be noted that of all algorithms included in benchmarking, $PCLRC_b$ is the only method specifically developed for the inference of metabolic networks, rather than gene regulatory networks [92].

The similarity and lack of statistical separation between top performing algorithms supports the prior observation that AUROC does not effectively distinguish between well-performing methods [74]. Within correlation-based and information theoretic groups (shown in Figures 4.1**a** and 4.1**b**), differences in performance become more pronounced as sample size increases. In Figure 4.1**d** however, performance differences become more pronounced as sample size decreases. Though little statistically significant difference was observed, this may suggest that AUROC more effectively distinguishes between 'good' methods at smaller sample sizes.

4.1.2 AUPR

Amongst correlation-based methods, AUPR indicates that $CORR_k$ is the top performing algorithm across small sample sizes (see Figure 4.2). Although $CORR_s$ produced lower average AUPR than $CORR_k$ across all sample sizes, no statistically significant difference was found between their AUPR. For large sample sizes of 100 and 200, the AUPR results closely resemble those based on AUROC (see Figure 4.1**a**), with *bicor* indicated as the top performing algorithm.

In contrast to the AUROC results observed for $PCLRC$ variations, which showed all algorithms receiving high AUROC scores (see Figure 4.1**b**), all $PCLRC$ variations are shown to perform poorly based on AUPR. The AUPR results shown in Figure 4.2**b** suggest that the performance of $PCLRC$ improves somewhat with decreasing sample size. In particular, $PCLRC_b$ was shown to statistically significantly outperform all other variations for small sample sizes 20 and 50. While the overall patterns and magnitude of scores observed across sample sizes differ between AUROC and AUPR, both measures indicate $PCLRC_b$ as the overall top performing algorithm of $PCLRC$ variations across small sample sizes.

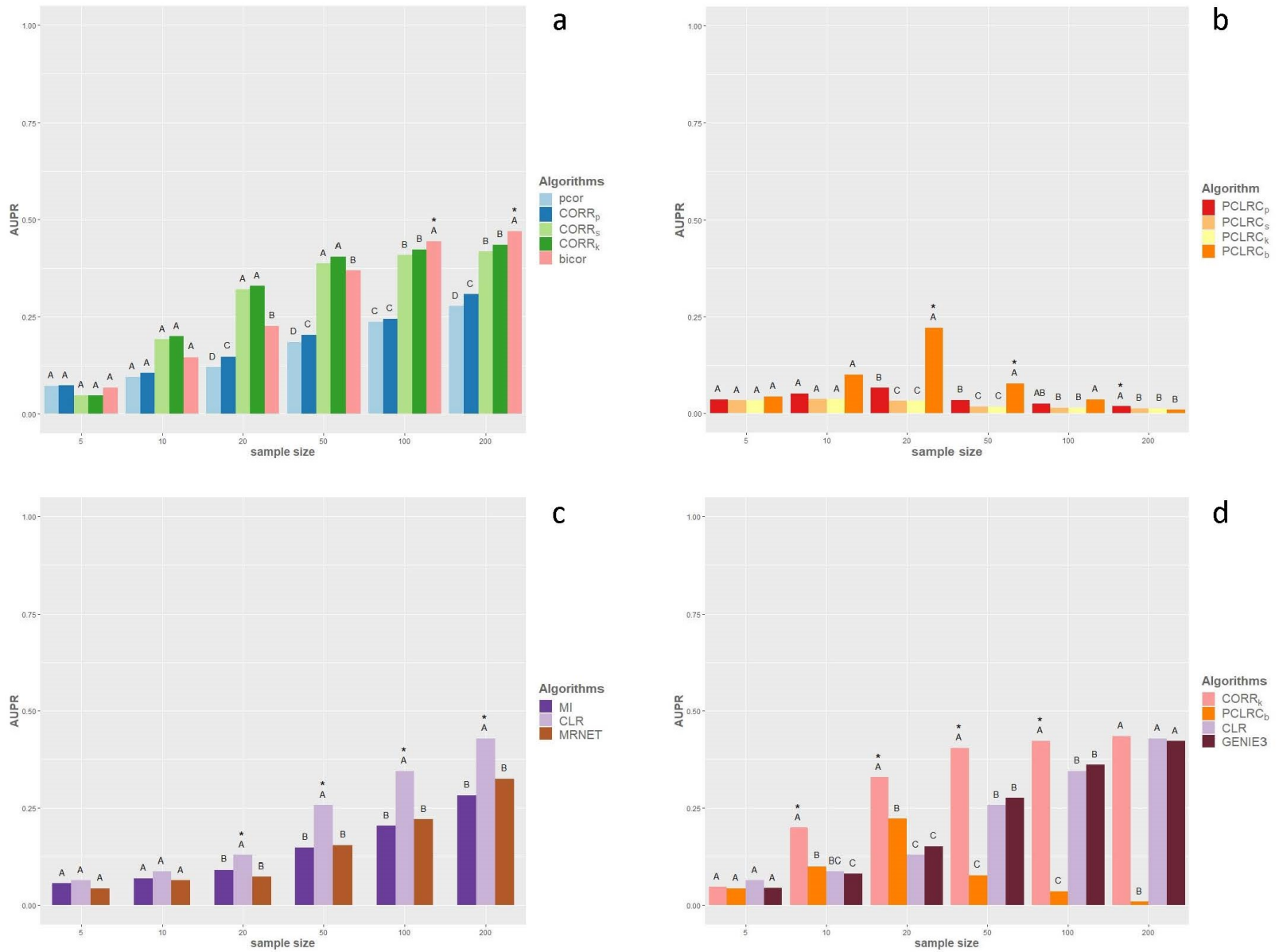

Figure 4.2: AUPR values averaged over 100 runs for various network inference methods. Performance was evaluated against simulated data with known structures for a range of sample sizes. Different letters denote a statistically significant difference ($p < 0.05$) among averages at each sample size according to a two-sided permutation test. Asterisks indicate methods which performed statistically significantly better than all other methods for a given sample size. Subfigure **a** compares AUPR values for correlation-based methods. Subfigure **b** compares AUPR values for *PCLRC* variations. Subfigure **c** compares AUPR values for Information Theoretic methods. Subfigure **d** compares AUPR values for the top performing methods from subfigures **a**–**c**

The AUPR results for information theoretic algorithms (see Figure 4.2**c**) closely resemble those based on AUROC (see Figure 4.1**c**), but indicate greater distinction between CLR and other algorithms. Based on AUPR, CLR is statistically significantly outperforms both MI and MRNET for and small sample sizes of 20 and 50, and large sample sizes of 100 and 200.

Directly comparing the top performing algorithms from each group (see Figure 4.2**d**) further highlights the difference in conclusions that can be drawn from AUROC and AUPR. AUPR scores clearly indicate $CORR_k$ as the top performing algorithm overall, while AUROC scores indicate almost no statistically significant differences across all sample sizes. These results suggest that AUPR may provide better discrimination between algorithm performances across small sample sizes.

4.1.3 MCC

In contrast to the AUROC results shown in Figure 4.1**a**, the MCC scores shown in Figure 4.3**a** indicate that $CORR_s$ and $CORR_k$ are the best performing correlation-based methods across all sample sizes. These results agree with those based on AUPR for small sample sizes of 10, 20, and 50 (see Figure 4.2**a**). Based on MCC, $CORR_k$ was found to statistically significantly outperform *pcor* and $CORR_p$ for small sample sizes of 10, 20, and 50, and large sample sizes of 100 and 200. $CORR_k$ also showed statistically significant improvement over *bicor* for sample sizes of 50, 100, and 200. $CORR_s$ statistically significantly outperformed $CORR_p$ and *pcor* for large sample sizes of 100 and 200, but not for small sample sizes. The statistically significant difference in performance observed between $CORR_s$ and $CORR_p$ contradicts previous results comparing the two algorithms based on AUROC and AUPR [48].

Similar to the AUROC and AUPR $PCLRC$ variation results (shown in Figures 4.1**b** and 4.2**b**, respectively) MCC scores indicate $PCLRC_b$ as the overall top performing algorithm. $PCLRC_b$ achieved statistically significantly greater MCC scores than all other $PCLRC$ variations for the small sample size of 50 and large sample sizes 100 and 200. In contrast to AUROC results, $PCLRC_p$ statistically significantly outperforms $PCLRC_s$ and $PCLRC_k$ for large sample sizes of 100 and 200. The pattern seen across $PCLRC$ variations for sample sizes 50, 100, and 200, is inverse to that observed for correlation methods. Specifically, the correlation methods showing the worst performance at those sample sizes achieved the best performance when used as the base similarity measure within the $PCLRC$ algorithm.

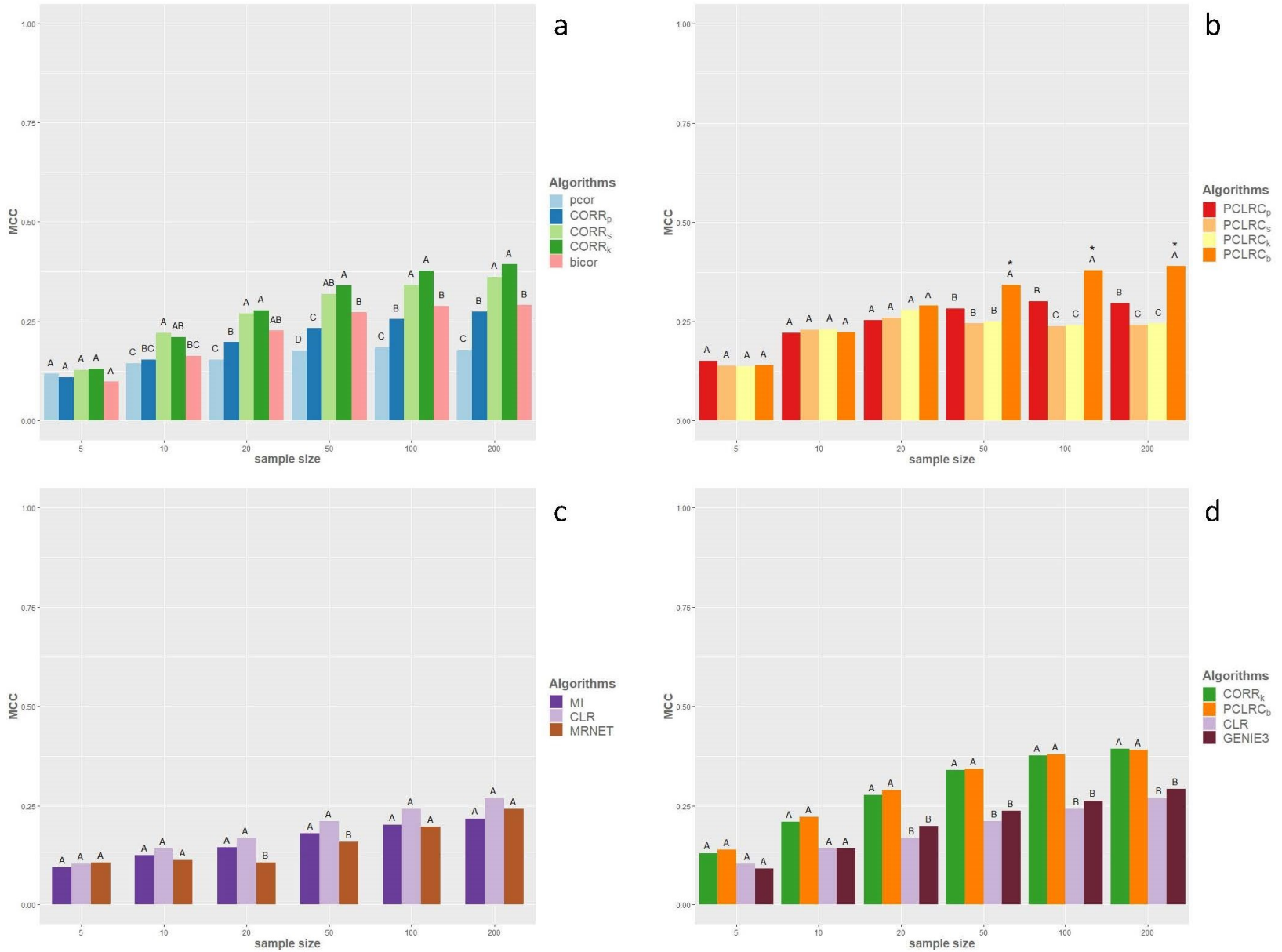

Figure 4.3: MCC values averaged over 100 runs for various network inference methods. Performance was evaluated against simulated data with known structures for a range of sample sizes. Different letters denote a statistically significant difference ($p < 0.05$) among averages at each sample size according to a two-sided permutation test. Asterisks indicate methods which performed statistically significantly better than all other methods for a given sample size. Subfigure **a** compares MCC values for correlation-based methods. Subfigure **b** compares MCC values for *PCLRC* variations. Subfigure **c** compares MCC values for Information Theoretic methods. Subfigure **d** compares MCC values for the top performing methods from Subfigures **a**-**c**

Similarities can be observed between the MCC (Figure 4.3**c**) and AUROC (Figure 4.1**c**) results for information theoretic algorithms. No statistically significant differences were found between any MCC scores for any sample size other than 20. However, CLR does achieve the overall highest average MCC score for sample sizes of 10 through 200.

Directly comparing top performing algorithms in Figure 4.3**d**, the differences in MCC scores contrasts the similarity of AUROC scores shown in Figure 4.1**d**. $PCLRC_b$ and $CORR_k$ statistically significantly outperformed both CLR and $GENIE3$ across small sample sizes of 20 and 50, and large sample sizes of 100 and 200. The MCC scores for $PCLRC_b$ and $CORR_k$ are comparable across all sample sizes.

4.1.4 FDR and PPV

For the goal of identifying direct network interactions, minimal FDR (and therefore maximal PPV) is desirable when comparing inferred networks to the primary interaction network, N_{ref} (see Section 2.1.1 for definitions of FDR and PPV). Generally, lower FDRs (shown in Figure 4.4) appear to correspond with higher MCC and AUROC scores. Although visualizations of the average PPVs for each algorithm are included for reference (see Figure 4.5), only FDRs will be discussed in detail, since FDR and PPV comprise parts of the same whole.

Differences between the average FDRs of correlation-based algorithms (shown in Figure 4.4**a**) align more closely with patterns observed in the corresponding MCC scores, rather than AUROC scores. $CORR_s$ and $CORR_k$ achieve statistically significantly lower FDRs for small sample sizes of 20 and 50 and large sample sizes of 100 and 200. $CORR_k$ additionally achieved a statistically significantly lower FDR than $CORR_s$ for a large sample size of 200.

Differences between the average FDRs for $PCLRC$ variations (shown in Figure 4.4**b**) similarly reflect differences observed between MCC scores (shown in Figure 4.3**b**). $PCLRC_p$ and $PCLRC_b$ achieve statistically significantly lower FDRs than $PCLRC_s$ and $PCLRC_k$ for the small sample size of 50, and large sample sizes of 100 and 200.

No statistically significant differences in the FDRs of different information theoretic methods were observed across any sample sizes (see Figure 4.4**c**). The lack of distinction between scores are consistent with those observed for the MCC and AUROC scores (see Figures 4.1**c** and 4.3**c**). However unlike $PCLRC$ and correlation-based

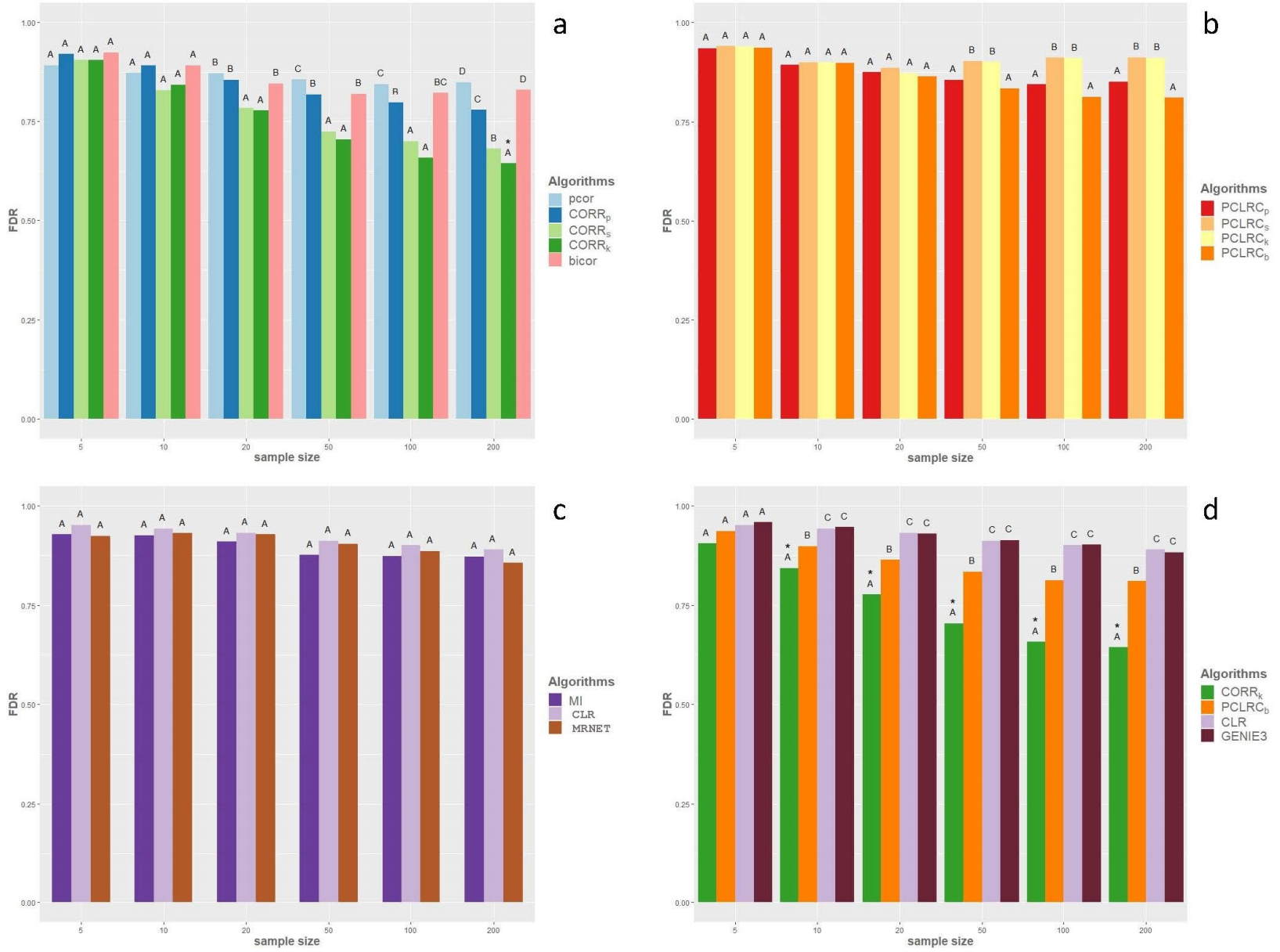

Figure 4.4: FDR values averaged over 100 runs for various network inference methods. Performance was evaluated against simulated data with known structures for a range of sample sizes. Different letters denote a statistically significant difference ($p < 0.05$) among averages at each sample size according to a two-sided permutation test. Asterisks indicate methods which performed statistically significantly better than all other methods for a given sample size. Subfigure **a** shows FDR values for correlation-based methods. Subfigure **b** shows FDR values for *PCLRC* variations. Subfigure **c** shows FDR values for Information Theoretic methods. Subfigure **d** shows FDR values for the top performing methods from subfigures **a**-**c**.

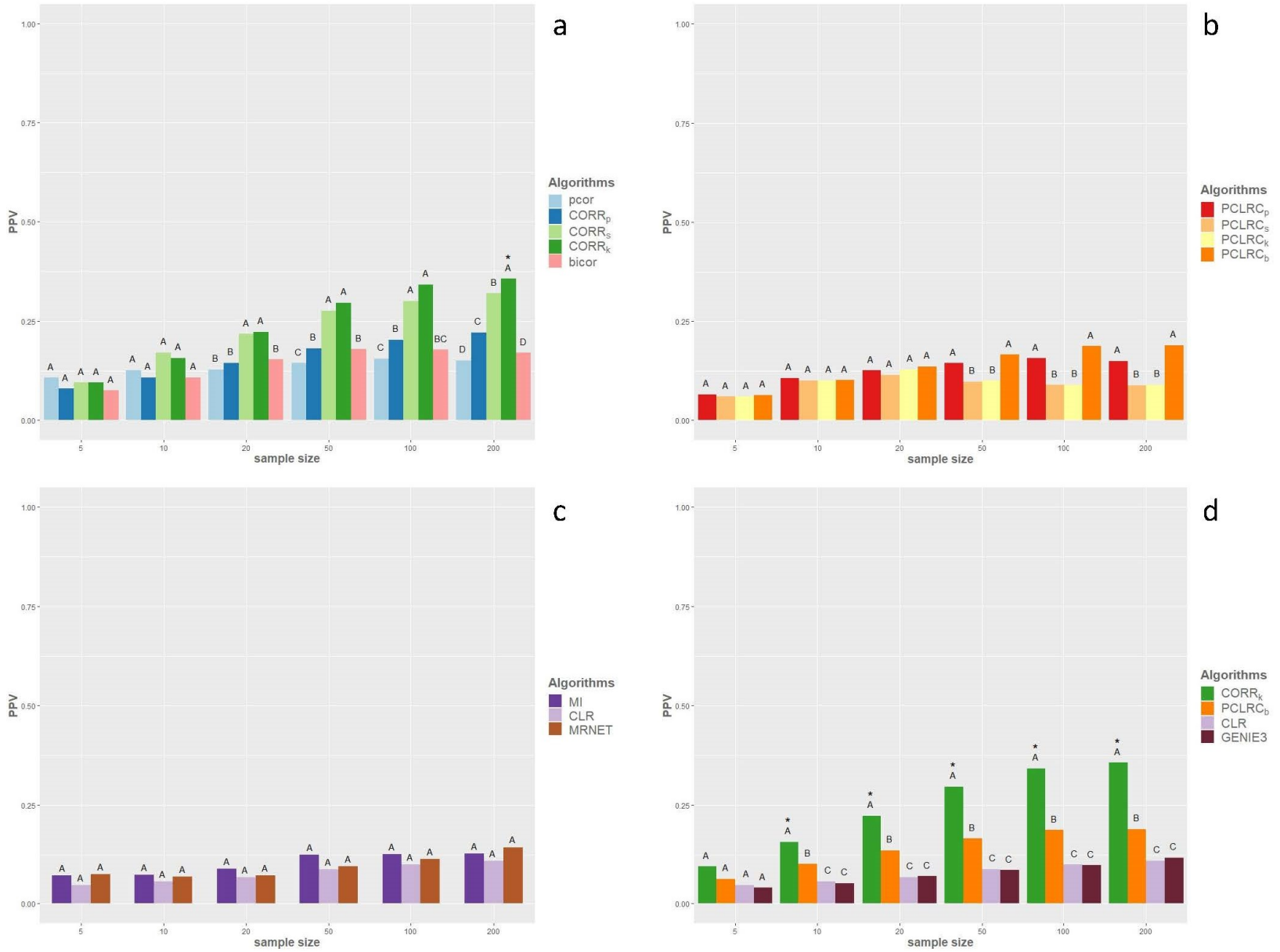

Figure 4.5: PPV values averaged over 100 runs for various network inference methods. Performance was evaluated against simulated data with known structures for a range of sample sizes. Different letters denote a statistically significant difference ($p < 0.05$) among averages at each sample size according to a two-sided permutation test. Asterisks indicate methods which performed statistically significantly better than all other methods for a given sample size. Subfigure **a** shows PPV values for correlation-based methods. Subfigure **b** shows PPV values for *PCLRC* variations. Subfigure **c** shows PPV values for Information Theoretic methods. Subfigure **d** shows PPV values for the top performing methods from subfigures **a-c**.

algorithms, CLR (which achieved the highest average MCC and AUROC scores for sample sizes 10 through 200) also produced the highest average FDR for those same sample sizes.

4.1.5 Comparison of Performance Measures

Across AUROC, AUPR, and MCC scores, the number of statistically significant differences observed decreases as sample sizes decreases, with no significant differences observed between any algorithms for the sample size of 5, and very few for the sample size of 10. Only correlation-based methods achieved statistically significant distinction in performance for sample size 10. The comparable performance of all algorithms in these cases suggest that sample sizes of 5 and 10 (approximately 0.06 and 0.12 times the number of variables in the simulated data set, respectively) may be insufficient for the prediction of statistically significant interactions in network reconstructions. However, since this observation can only be made for the set number of variables included in the simulated data set, additional benchmarking against a wider variety of data set dimensions would need to be considered to validate this hypothesis.

4.2 Structural Analyses

In network analysis, different measures of centrality can be used to answer questions of biological and/or clinical significance (see Section 2.2.1). It is therefore important to investigate the effect that different network inference algorithms have on the conclusions that can be drawn from such analyses. In the following section I present the results of several node centrality analyses performed on networks predicted using the top performing algorithms, as determined in Section 4.1.

Based on AUROC, AUPR, and MCC measures, $PCLRC_b$ and $CORR_k$ were determined to be the overall top performing algorithms for the inference of metabolic association networks from small sample sizes. Weighted $CORR_k$ and $PCLRC_b$ networks, upon which centrality analyses can be performed, were predicted from both simulated and experimental data. For networks inferred from simulated data, different node centrality measures for the AA reference network (N_{ref}), $CORR_k$ predicted network (N_k), and $PCLRC_b$ predicted network (N_b) are compared in Figures 4.7–4.9 in Section 4.2.1. In addition to the aforementioned analyses performed on networks predicted from simulated data, DCA was performed on experimental data. The effects

of $PCLRC_b$ and $CORR_k$ on DCA results are discussed in Section 4.2.2 to highlight the effects that choice of network inference algorithm can have on biomarker identification, disease diagnostics, and metabolic regulation studies.

4.2.1 Centrality Measures

The reference AA metabolism network (N_{ref}) consists of 83 nodes and 83 edges (shown in Figure 4.6). Predicting N_{ref} from simulated data, $CORR_k$ produced a network N_k consisting of 71 nodes, indicating that 12 metabolites were not predicted to interact with any others, and 425 edges (shown in subfigure **a** of Figures 4.7 - 4.9). Predicting N_{ref} from the same data, $PCLRC_b$ produced a network N_b consisting of 83 nodes and 503 edges (shown in subfigure **c** of Figures 4.7 - 4.9). Note that while $CORR_k$ achieves fewer FPs than $PCLRC_b$, $CORR_k$ did not predict $X.AA$, an undeniably crucial player in AA metabolism, to interact with any other metabolite. The large number of FPs present in N_b obscure any visually identifiable community structures. In contrast, within N_k (containing notably fewer FPs) two distinct clusters are observable.

Within the AA metabolism reference network (N_{ref}), the greatest betweenness centralities are belong to nodes $X.LTA4$, $X.15\_HpETE$, $X.AA$, and $X.PGH2$. This can be observed in Figure 4.6**b** and Table 4.1. It can be visually observed that each of these nodes are the only intermediate point of access for a large number of other low-degree nodes. When looking at the corresponding networks predicted using $PCLRC_b$ and $CORR_k$, this structure is not well defined. Comparing subfigures 4.7**d** and 4.6**b**, while the edges of N_{ref} appear to be correctly predicted, because of the large number of FPs, the betweenness of nodes was incorrectly captured in N_b. The more distinct structural characteristics of N_k result in greater variation in betweenness of the network's nodes. However, comparing subfigures 4.7**b** and 4.6**b**, this variation does not reflect improvement over $PCLRC_b$ in the reconstruction of betweenness centralities observed in N_{ref}. Looking at the nodes in each network with the highest betweenness centralities (shown in Table 4.1), seven results are shared between N_k and N_{ref}, while eight are shared between N_b and N_{ref}. Three results are additionally shared only between N_k and N_b.

A more distinct difference in the results produced from the $CORR_k$ and $PCLRC_b$ networks can be observed for closeness centrality (as shown in Figure 4.8). Through visual comparison of Figures 4.7 and 4.6**c**, $PCLRC_b$ appears to more closely reconstruct the closeness patterns observed in N_{ref}. The notably decreased closeness

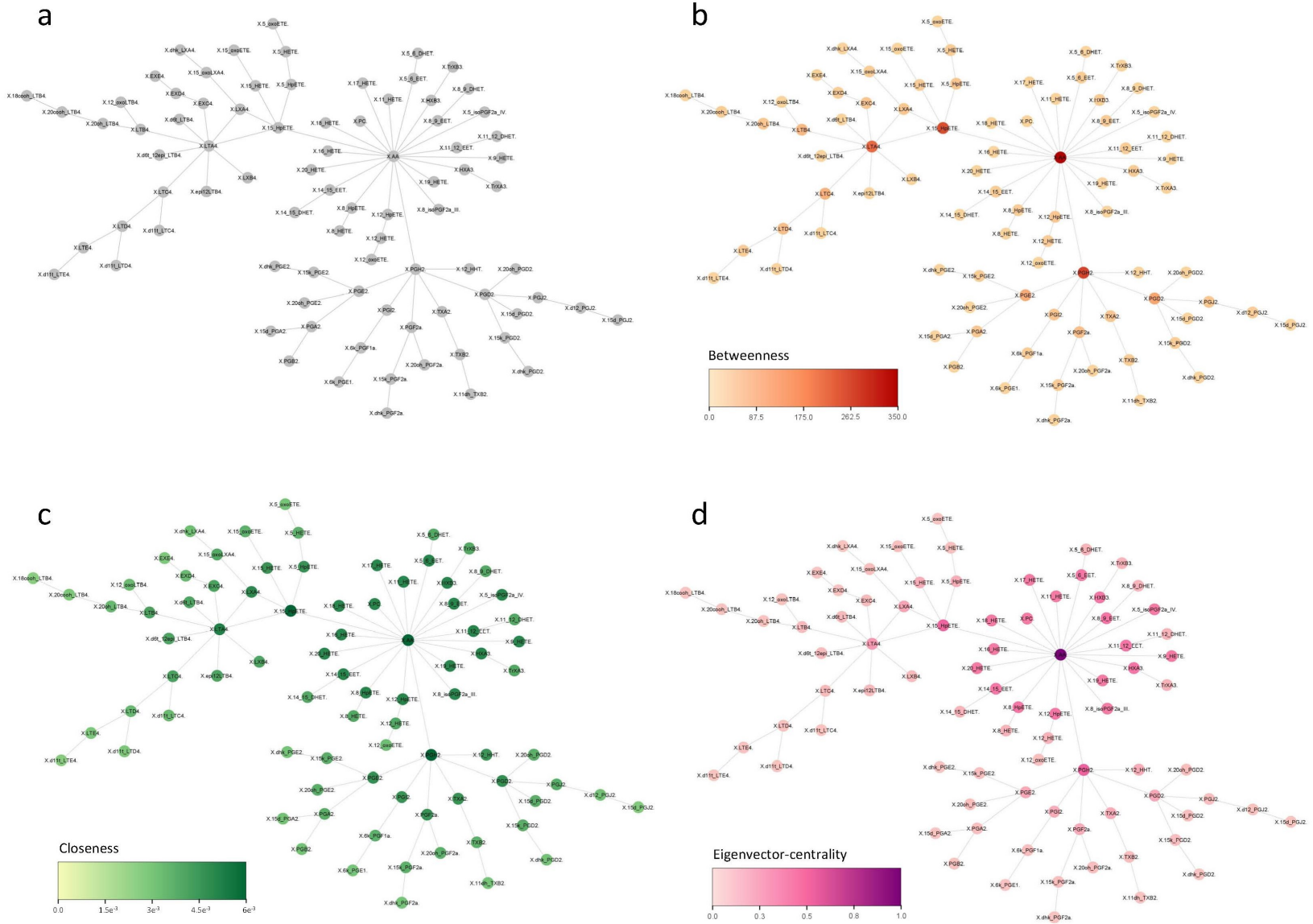

Figure 4.6: Structural characteristics of the AA metabolism reference network ($\mathbf{N}_{ref}$). Subfigure **a** depicts ($\mathbf{N}_{ref}$). Nodes are labeled with the corresponding metabolite names. Lines between nodes indicate a direct interaction between metabolites. Subfigure **b** shows the betweenness scores calculated for each node in the network. Betweenness increases with darkening node color, as indicated by the color scale below subfigure **b**. Subfigure **c** shows the closeness scores calculated for each node in the network. Closeness increases with darkening node color, as indicated by the color scale below subfigure **c**. Subfigure **b** shows the eigenvector-centrality scores calculated for each node in the network. Eigenvector-centrality increases with darkening node color, as indicated by the color scale below subfigure **d**.

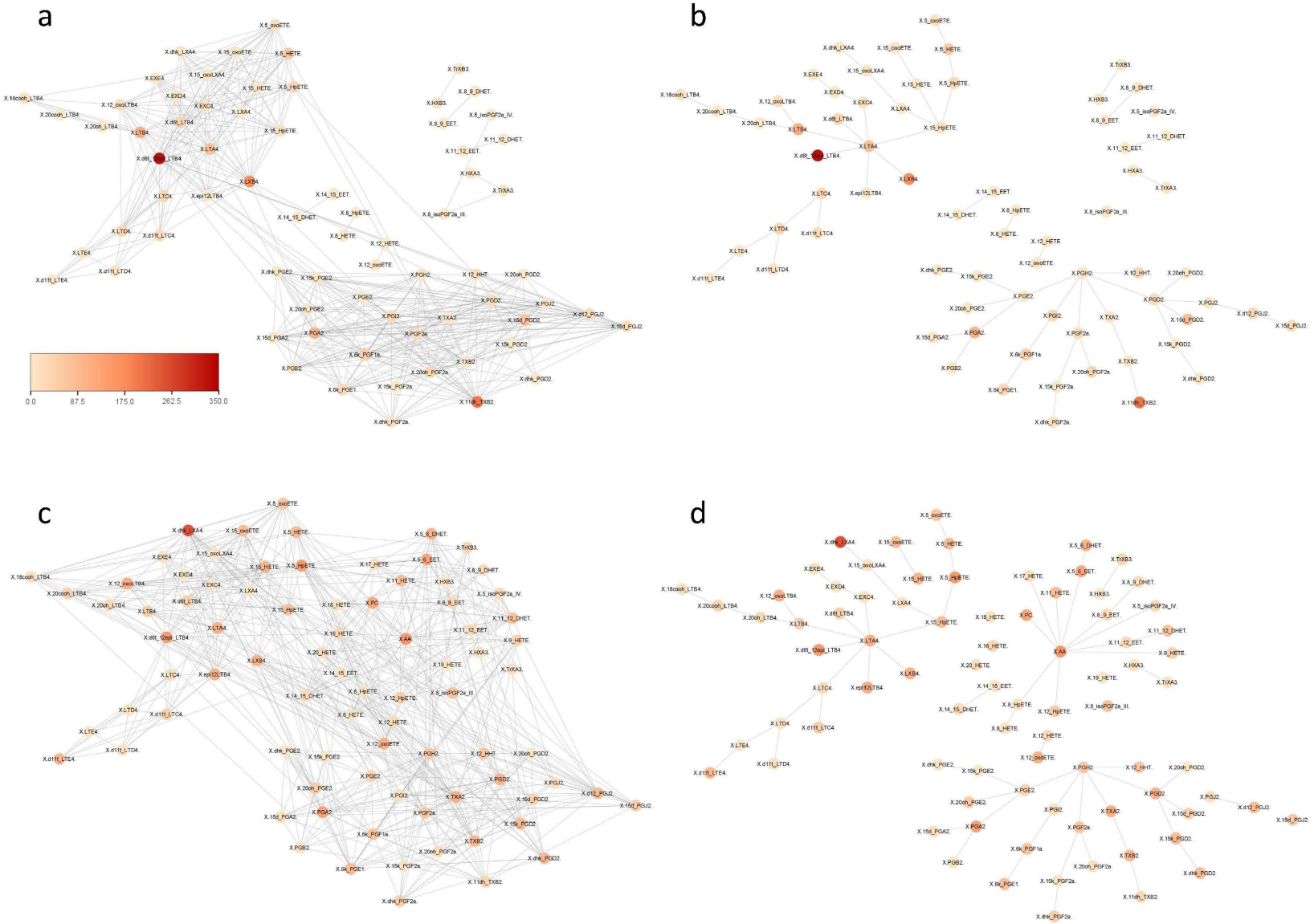

Figure 4.7: Betweenness centrality within networks predicted from simulated data using different network inference methods. In each network, nodes are colored by betweenness centrality score and labeled with the corresponding metabolite names. Lines between nodes indicate a predicted interaction between metabolites. Subfigure **a** depicts the network predicted using $CORR_k$, N_k. Subfigure **b** shows the intersection of N_{ref} and N_k. Subfigure **c** depicts the network predicted using $PCLRC_b$, N_b. Subfigure **d** shows the intersection of N_{ref} and N_b.

Table 4.1: The top 20 nodes with the highest betweenness centralities for the AA metabolism reference network N_{ref}, N_k, N_b. Cells highlighted in green are those shared between each algorithm and N_{ref}. Cells highlighted in yellow indicate results shared only between N_k and N_b.

N_{ref}	N_k	N_b
X.AA.	X.d6t_12epi_LTB4.	X.dhk_LXA4.
X.PGH2.	X.11dh_TXB2.	X.AA.
X.15_HpETE.	X.LXB4.	X.d6t_12epi_LTB4.
X.LTA4.	X.LTB4.	X.5_HpETE.
X.PGD2.	X.PGA2.	X.PGA2.
X.PGE2.	X.15d_PGD2.	X.PC.
X.LTC4.	X.LTA4.	X.5_6_EET.
X.LTB4.	X.5_HETE.	X.12_oxoETE.
X.LTD4.	X.6k_PGF1a.	X.12_oxoLTB4.
X.PGF2a.	X.20oh_PGF2a.	X.epi12LTB4.
X.PGA2.	X.d6t_LTB4.	X.LTA4.
X.TXA2.	X.LTD4.	X.TXA2.
X.PGI2.	X.5_HpETE.	X.PGD2.
X.PGJ2.	X.LTC4.	X.LXB4.
X.LXA4.	X.PGI2.	X.15_HETE.
X.5_HpETE.	X.6k_PGE1.	X.5_6_DHET.
X.20oh_LTB4.	X.d12_PGJ2.	X.TXB2.
X.EXC4.	X.d11t_LTC4.	X.d11t_LTE4.
X.12_HpETE.	X.15d_PGJ2.	X.5_HETE.
X.TXB2.	X.PGD2.	X.dhk_PGD2.

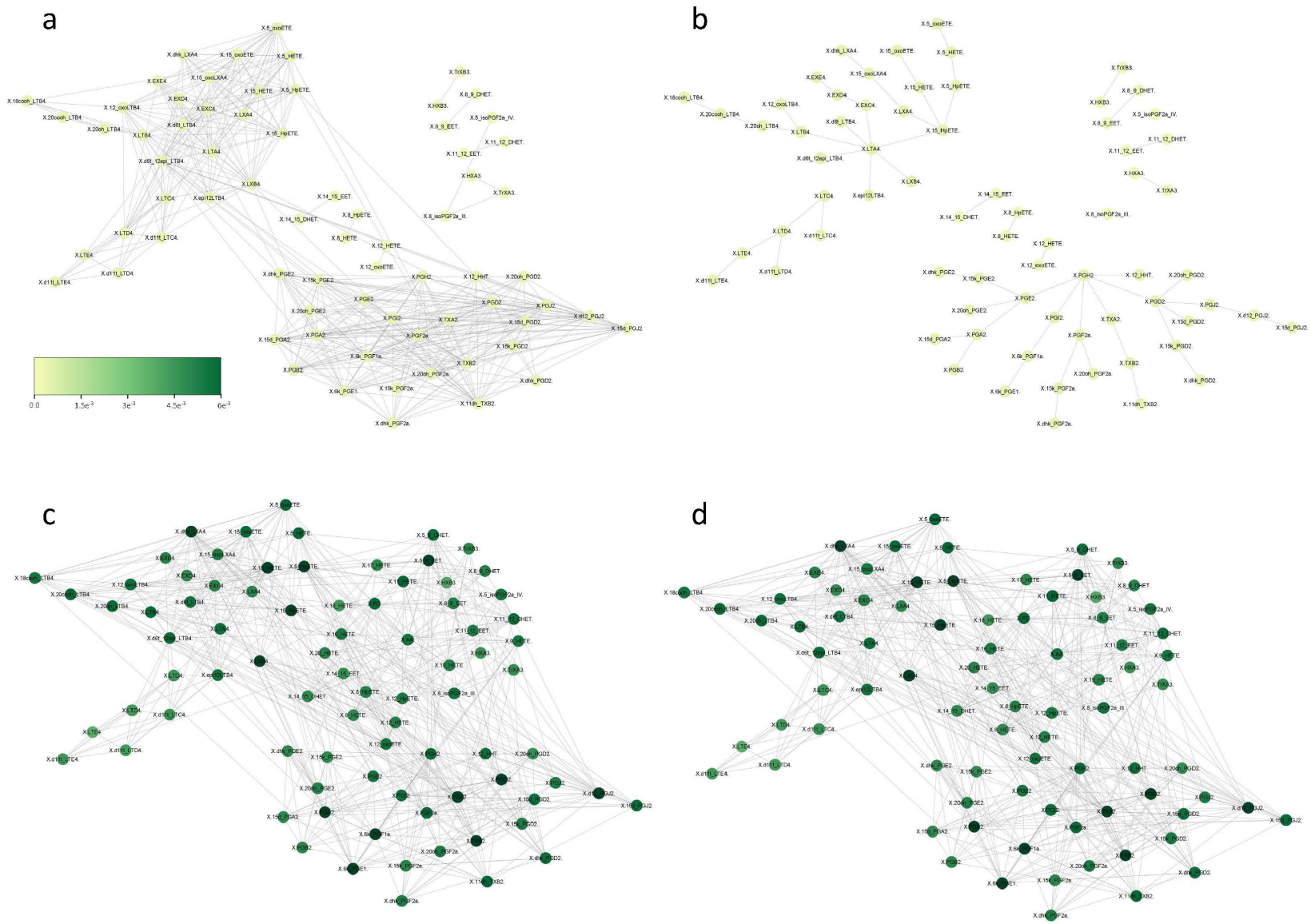

Figure 4.8: Closeness centrality within networks predicted from simulated data using different network inference methods. In each network, nodes are colored by closeness centrality score and labeled with the corresponding metabolite names. Lines between nodes indicate a predicted interaction between metabolites. Subfigure **a** depicts the network predicted using $CORR_k$, N_k. Subfigure **b** shows the intersection of N_{ref} and N_k. Subfigure **c** depicts the network predicted using $PCLRC_b$, N_b. Subfigure **d** shows the intersection of N_{ref} and N_b.

Table 4.2: The top 20 nodes with the highest closeness centralities for the AA metabolism reference network N_{ref}, N_k, N_b. Cells highlighted in green are those shared between each algorithm and N_{ref}. Cells highlighted in yellow indicate results shared only between N_k and N_b.

N_{ref}	N_k	N_b
X.AA.	X.d6t_12epi_LTB4.	X.dhk_LXA4.
X.15_HpETE.	X.11dh_TXB2.	X.5_HpETE.
X.PGH2.	X.LXB4.	X.LXB4.
X.LTA4.	X.PGA2.	X.6k_PGE1.
X.12_HpETE.	X.LTA4.	X.PGA2.
X.8_HpETE.	X.6k_PGF1a.	X.TXB2.
X.HXA3.	X.PGI2.	X.TXA2.
X.HXB3.	X.6k_PGE1.	X.15_HpETE.
X.5_6_EET.	X.15d_PGJ2.	X.d12_PGJ2.
X.8_9_EET.	X.5_HETE.	X.15_HETE.
X.11_12_EET.	X.d12_PGJ2.	X.PGD2.
X.14_15_EET.	X.5_HpETE.	X.5_6_EET.
X.PC.	X.20oh_PGF2a.	X.6k_PGF1a.
X.11_HETE.	X.15d_PGD2.	X.11_HETE.
X.9_HETE.	X.LTB4.	X.PC.
X.16_HETE.	X.12_HHT.	X.20cooh_LTB4.
X.17_HETE.	X.EXC4.	X.PGF2a.
X.18_HETE.	X.d6t_LTB4.	X.11dh_TXB2.
X.19_HETE.	X.15_HpETE.	X.5_HETE.
X.20_HETE.	X.12_oxoLTB4.	X.5_oxoETE.

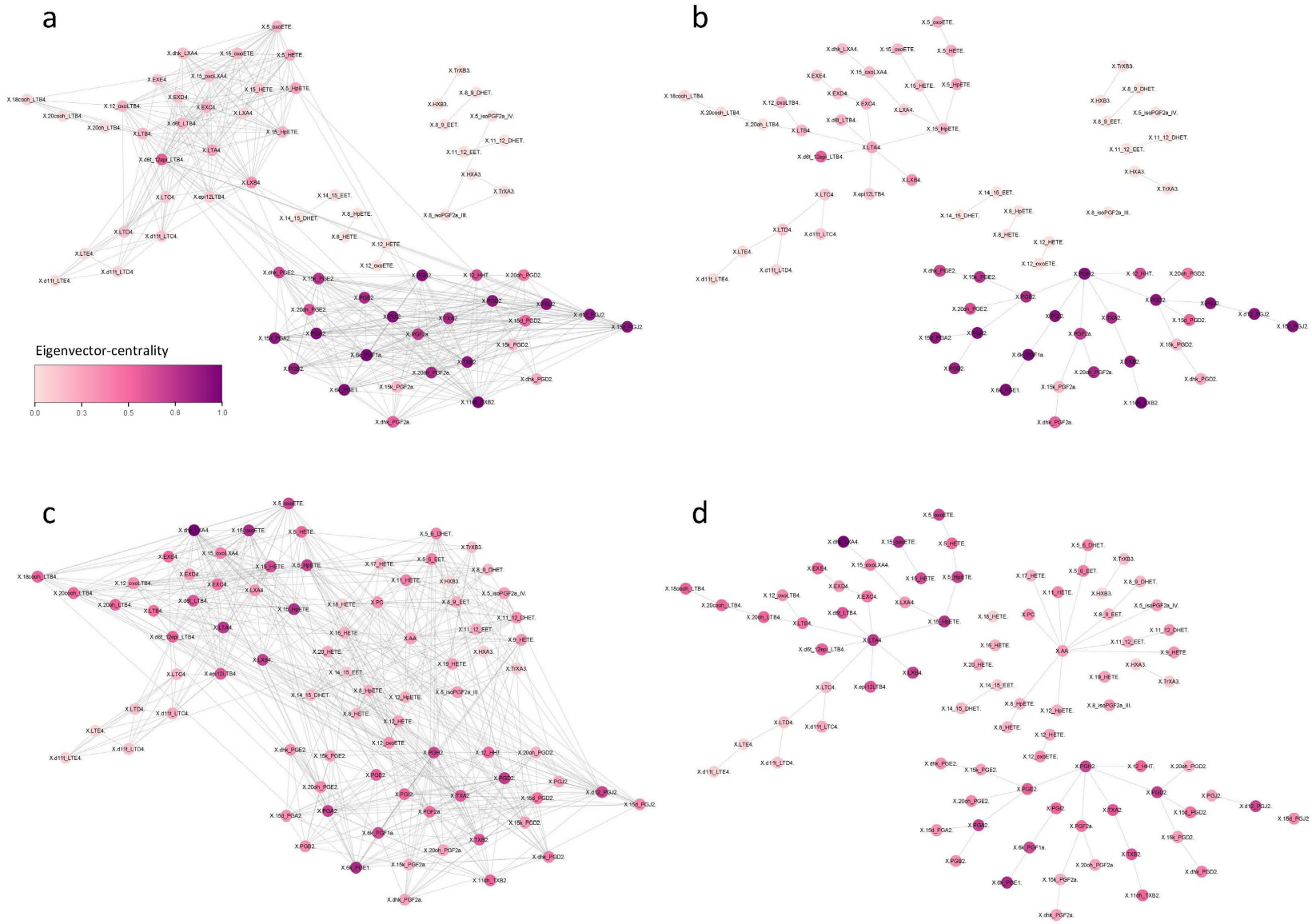

Figure 4.9: Eigenvector centrality within networks predicted from simulated data using different network inference methods. In each network, nodes are colored by eigenvector centrality score and labeled with the corresponding metabolite names. Lines between nodes indicate a predicted interaction between metabolites. Subfigure **a** depicts the network predicted using $CORR_k$, N_k. Subfigure **b** shows the intersection of N_{ref} and N_k. Subfigure **c** depicts the network predicted using $PCLRC_b$, N_b. Subfigure **d** shows the intersection of N_{ref} and N_b.

Table 4.3: The top 20 nodes with the highest eigenvector centralities for the AA metabolism reference network N_{ref}, N_k, N_b. Cells highlighted in green are those shared between each algorithm and N_{ref}. Cells highlighted in yellow indicate results shared only between N_k and N_b.

N_{ref}	N_k	N_b
X.AA.	X.11dh_TXB2.	X.dhk_LXA4.
X.PGH2.	X.6k_PGF1a.	X.6k_PGE1.
X.15_HpETE.	X.PGA2.	X.15_oxoETE.
X.12_HpETE.	X.6k_PGE1.	X.15_HpETE.
X.8_HpETE.	X.PGI2.	X.LTA4.
X.HXA3.	X.PGH2.	X.PGD2.
X.HXB3.	X.15d_PGJ2.	X.5_HpETE.
X.5_6_EET.	X.PGJ2.	X.LXB4.
X.8_9_EET.	X.d12_PGJ2.	X.PGA2.
X.11_12_EET.	X.TXB2.	X.5_oxoETE.
X.14_15_EET.	X.PGD2.	X.d12_PGJ2.
X.PC.	X.15d_PGA2.	X.PGH2.
X.11_HETE.	X.PGB2.	X.6k_PGF1a.
X.9_HETE.	X.TXA2.	X.15_HETE.
X.16_HETE.	X.20oh_PGF2a.	X.TXB2.
X.17_HETE.	X.PGE2.	X.TXA2.
X.19_HETE.	X.15k_PGE2.	X.d6t_LTB4.
X.8_isoPGF2a_III.	X.PGF2a.	X.epi12LTB4.
X.5_isoPGF2a_IV.	X.dhk_PGE2.	X.PGE2.
X.18_HETE.	X.12_HHT.	X.d6t_12epi_LTB4.

centralities in N_k is likely due to its lower edge density and distinct clusters (see Figure 4.8**a**).

Though the apparent similarity in closeness magnitudes between N_b and N_{ref} networks may suggest superior reconstruction using the $PCLRC_b$ algorithm, this result is not reflected in Table 4.2. Of the 20 N_{ref} nodes with the greatest closeness centralities, only two appear in the top 20 nodes of N_k, and four in N_b. The $CORR_k$ and $PCLRC_b$ networks share a greater number of top 20 nodes (eight) with each other than either do with N_{ref}. These results indicate that while the reduced density of N_k produced distinctly lower closeness centralities than seen in N_b, the overall structure of closeness centralities are most similar between the N_b and N_k networks. Therefore, $PCLRC_b$ does not better capture closeness centrality than $CORR_k$.

Based on the eigenvector centralities visualized within in N_k and N_b in Figure 4.9, both networks induced larger eigenvector centralities across more nodes than those observed in N_{ref}. Of the nodes found to have the greatest eigenvector centralities in N_{ref}, only one was commonly identified in N_k and two in N_b (shown in Table 4.3). Based on these results, both algorithms failed to identify the majority of highly influential nodes of the AA metabolism network. However, nine nodes were commonly identified as highly influential between N_k and N_b. Similar to betweenness and closeness centrality results, greater similarity can be observed between the predicted networks than between either of the predicted networks and the reference N_{ref}.

Identifying nodes of the greatest betweenness, closeness, or eigenvector centralities is one way to identify metabolites of different functional importance within a network. Unexpectedly, the difference in network density observed between N_k and N_b had the greatest effect on closeness, followed by eigenvector centrality, and lastly, betweenness. Although distinct differences in the centrality patterns of N_k and N_b could be visually identified, these differences had little consequence on the identification of nodes with the greatest centrality measures. Across all included measures, $CORR_k$ and $PCLRC_b$ captured a similar number of nodes with the greatest centralities in N_{ref}. Based on the results of these structural analyses, neither algorithm presents a distinct advantage over the other for the inference of metabolic samples from a small sample size. However, the exclusion of nodes (metabolites present in the data) observed in the N_k could significantly impact a variety of network analyses. I would therefore recommend use of $PCLRC_b$ over $CORR_k$ for this application, despite the increased FDR.

The lack of distinction between the ability of the $CORR_k$ and $PCLRC_b$ algo-

rithms to capture centrality structure within N_{ref} reflects the similarity observed between MCC scores (in contrast to AUROC and AUPR measures). Contradictory to these results, AUROC indicated statistically significantly improved performance of $PCLRC_b$ over $CORR_k$ (see Figure 4.1) for sample size 20, while AUPR indicated the reverse. This suggests that where AUROC suffers from bias towards the reference rather than the predicted network (over-rewarding TPs and under-penalizing FPs), AUPR suffers from an opposite bias (over-rewarding TNs and under-penalizing penalizing FNs). I therefore would recommend use of MCC over AUROC and AUPR for a more balanced assessment of network prediction performance.

4.2.2 Experimental Data and DCA

In addition to centrality measures, I chose to investigate the effects of $CORR_k$ and $PCLRC_b$ on the results of DCA, a comparative network analysis commonly applied to the study of human diseases (see Section 2.2.2). To perform DCA, networks needed to be predicted for two distinct disease states. Four networks were thus predicted using both $CORR_k$ and $PCLRC_b$ from two data sets containing the same set of metabolites; one containing lipid panels sampled from fasted Tg mice (fTg) and the other containing lipid panels sampled from fasted non-Tg mice ($fnTg$) (see Section 3.1.2). The resulting networks are shown as binary heatmaps in Figure 4.10.

Similar to networks predicted from simulated data, the networks predicted by $CORR_k$ from both non-Tg and Tg data are sparser than those predicted by $PCLRC_b$. The $CORR_k$ non-Tg and Tg networks consist of 1253 and 1195 edges, respectively. The $PCLRC_b$ non-Tg and Tg networks consist of 4026 and 4011 edges, respectively. DCA was performed on the $CORR_k$ and $PCLRC_b$ networks separately (as described in Section 3.2.3), the results of which are shown in Figure 4.11.

In contrast to the similarity in centrality structures observed between N_k and N_b, a stark difference can be observed in the DCA results. Based on networks predicted using $CORR_k$, 189 metabolites (90% of the 199 present in the networks) were determined to be statistically significantly differentially connected between Tg and non-Tg network states. Less than half as many metabolites (80 out of 199) were determined to be statistically significantly deferentially connected between networks predicted using $PCLRC_b$. The DCA results produced with the $CORR_k$ networks would provide no meaningful insights to differential centrality-based analyses, and the lack on *non*-differentially connected metabolites render the results practically

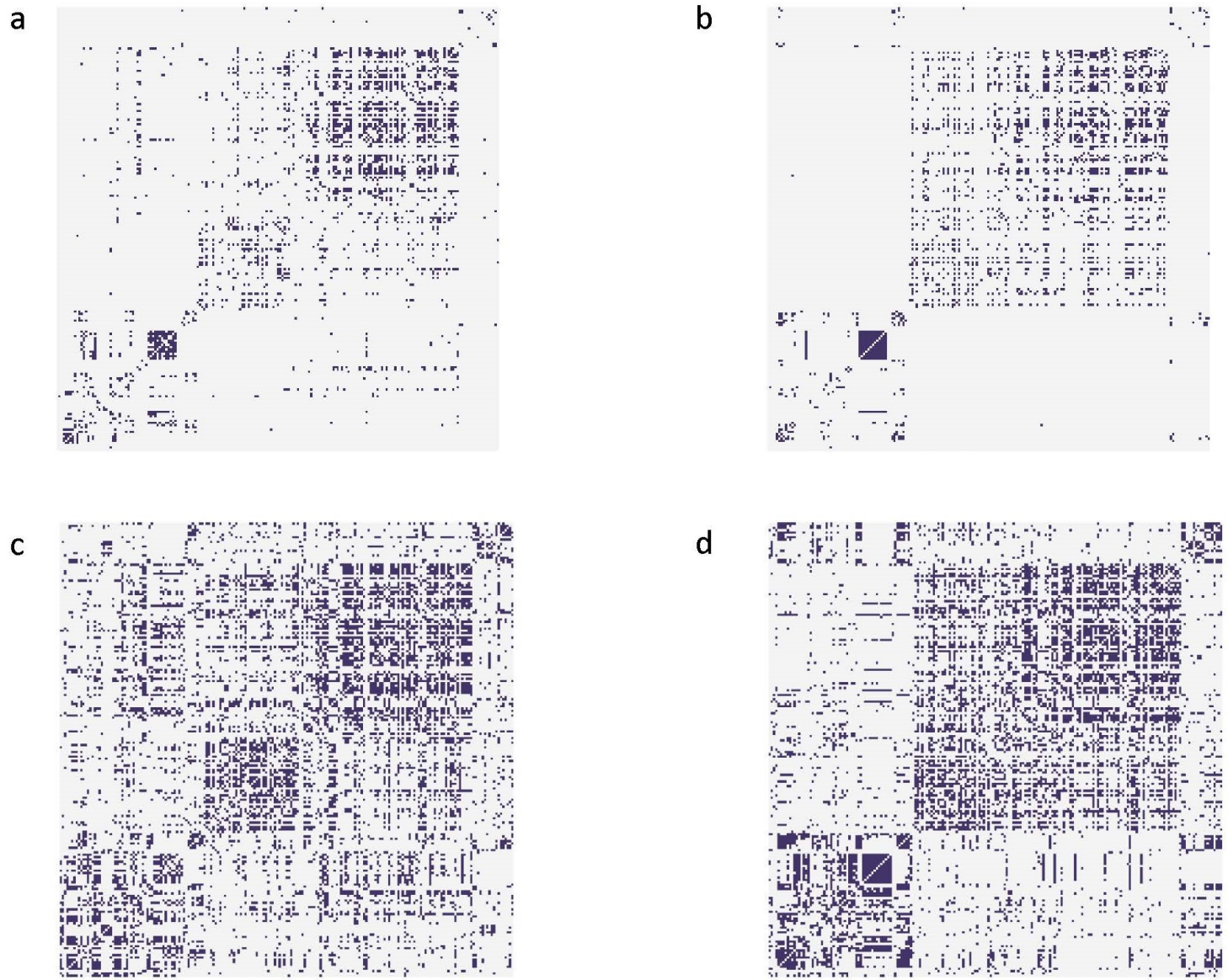

Figure 4.10: Networks inferred from experimental mouse-model Alzheimer's data using $CORR_k$ and $PCLRC_b$. Subfigures **a** and **c** show binary heatmap representations of networks inferred from $fnTg$ using $CORR_k$ and $PCLRC_b$, respectively. Subfigures **b** and **d** show binary heatmap representations of networks inferred from fTg using $CORR_k$ and $PCLRC_b$, respectively. For each heatmap, dark blue cells indicate a edge, while light grey cells indicate no edge.

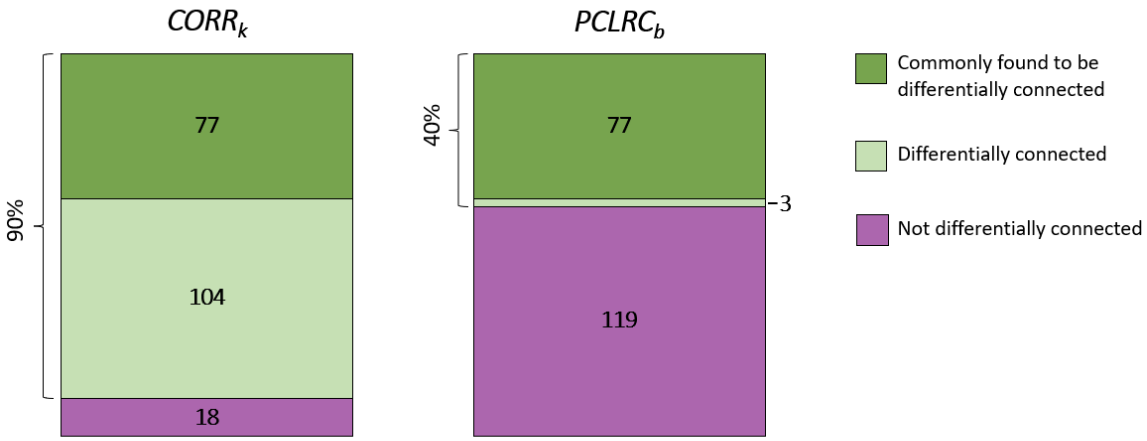

Figure 4.11: Abstract representation of the results of DCA performed on networks inferred using both $CORR_k$ and $PCLRC_b$ from Alzheimer's data. Each bar illustrates the proportion of network metabolites found to be differentially connected between Tg and non-Tg states for each algorithm. As shown in the $CORR_k$ results (left), 90% of all nodes in the networks inferred using $CORR_k$ were found to be differentially connected between states. As shown in the $PCLRC_b$ results (right), 40% of all network nodes were found to be differentially connected between states.

useless in biomarker identification studies. Although for small sample sizes $CORR_k$ was shown to outperform $PCLRC_b$ based on AUPR, and perform comparably based on MCC, $PCLRC_b$ has proven to provide more interpretable DCA results; a critical need for the study of disease state networks. As such, I recommend $PCLRC_b$ as the most suitable algorithm for the inference of small sample size metabolic networks for the study of human diseases.

Chapter 5

Conclusions and Future Work

Since metabolism is largely conserved between mice and humans, transgenic mouse models have become an increasingly common model from which to study AD [90]. Studies utilizing such models have significantly aided our understanding of the mechanisms of AD [3, 115] and the development of therapeutics. However, previous work has suggested that animal studies of neurological disorders are often underpowered as a result of small sample size and a large number of variables, among other factors [26, 97, 110]. Underpowered studies are at greater risk for false positive findings. It is therefore of great importance to identify network inference algorithms that perform high quality network reconstruction given small sample sizes (of fewer samples than variables).

Working towards improving the quality of network construction from small cohort disease studies, the aims of this thesis were threefold. First, to assess the performance of several popular network inference algorithms using multiple performance measures across a range of sample sizes, and identify top performing methods for small sample sizes. Cross-referencing AUROC, AUPR, MCC, and FDR benchmarking results in Section 4.1, I identified $CORR_k$ and $PCLRC_b$ as the overall top performing algorithms.

The second aim was to investigate the effects of different network inference algorithms (determined to perform comparably well in the aforementioned benchmarking) on different node centrality measures. Comparing the betweenness, closeness, and eigenvector centralities within the predicted $CORR_k$, $PCLRC_b$ networks against the AA metabolism reference network (in Section 4.2.1), I found no distinction between the effects of either algorithm. For all three centralities; there was greater similarity between the patterns observed in the two predicted networks than either predicted

network and the reference. However, based on the noteworthy difference DCA results observed for $CORR_k$ and $PCLRC_b$ Alzheimer's disease state networks and implicated effects on disease biomarker or diagnostic studies, I identified $PCLRC_b$ as the preferred metabolic network inference algorithm to be used in small sample size differential metabolic studies.

While $PCLRC_b$ reconstructed a greater proportion of the reference network than $CORR_k$, and indicated more discriminatory DCA results, the two algorithms performed comparably in the reconstruction of the reference network's structural properties. The noted differences in centrality structures between the $PCLRC_b$ network and N_{ref} (see Section 4.2) call into question whether the results presented by any centrality-based differential network analysis would accurately reflect structural differences between Tg and non-Tg states. In multiple studies, DCA has been applied to networks inferred using $PCLRC_p$ from metabolic data to identify molecular determinants associated with different health states [111, 112, 65]. Each of these studies, investigating blood pressure and heart rate [112], Acute Ischemic Stroke mortality [65], and Acute Myocardial Infarction mortality [111], indicated specific metabolites supported by existing literature as metabolic determinants. Due to the large proportion of metabolites observed to be deferentially connected in the $PCLRC_b$ networks, it is likely that some number of structural differences with biological relevancy are captured, as indicated by the aforementioned studies. For metabolites that lack supporting literature to elucidate structural differences observed in DCA, possibly indicating newfound mechanisms, it is critical that structural properties are accurately captured. This indicates the need for the development of a network inference algorithm that captures the majority of true network edges, like PCLRC, as well as network structure.

The final aim of this thesis was to identify the network inference performance measure best suited for the evaluation of metabolic network reconstruction from small sample sizes. Based on several observations, I identified MCC as the most suitable measure. While AUPR clearly indicated $CORR_k$ as the top performing methods, DCA results suggested poor performance of $CORR_k$ for the analysis of disease state network. DCA results along with the absence of critical nodes in the $CORR_k$ predicted AA metabolism network also suggest a bias in the AUPR measure of over-rewarding TNs and under-penalizing FNs. In comparison to AUROC (previously shown to be a problematic measure when applied to imbalanced data [74]), MCC demonstrated better discrimination between the performance of algorithms at small

sample sizes, and a greater correspondence between large MCC and small FDR. While I recommend MCC as the best existing performance measure for the assessment of small sample size metabolic studies, the results of this study indicate the need for a summative performance measure that also considered the structural properties of predicted networks.

One limitation of the presented study is the lack of diversity of data included in benchmarking and structural analyses. This study considered a range of sample sizes for a set number of variables. In the interest of establishing a recommended sample size for metabolic studies relative to the number of included metabolites, network inference algorithm performance should be re-assessed for a wider range of data set dimensions. Additionally, the observed effects of network inference algorithms on network structure may be relevant to gene regulatory networks as well as metabolic association networks. Wolf et al. determined node degree to be one of the most relevant topological feature for highlighting the mechanisms of gene regulatory networks [119]. It would therefore be of interest, to the advancement of small sample size gene regulatory network reconstruction, to repeat the presented study using gene expression data.

Bibliography

[1] Scott Ayton, Peng Lei, and Ashley I. Bush. Metallostasis in alzheimer's disease. *Free Radical Biology and Medicine*, 62:76–89, 2013.

[2] Kunihiro Baba, Ritei Shibata, and Masaaki Sibuya. Partial correlation and conditional correlation as measures of conditional independence. *Australian New Zealand Journal of Statistics*, 46(4):657–664, 2004.

[3] Kelly Bales. The value and limitations of transgenic mouse models used in drug discovery for alzheimer's disease: an update. *Expert Opinion on Drug Discovery*, 7(4):281–297, 2012.

[4] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 1999.

[5] Nir Barzilai, Derek M. Huffman, Radhika H. Muzumdar, and Andrzej Bartke. The critical role of metabolic pathways in aging. *Diabetes*, 61(6):1315–1322, 2012.

[6] Phillip Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.

[7] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PLOS ONE*, 12(6):e0177678, 2017.

[8] Andrew Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1997.

[9] Matthew Buczynski, Darren Dumlao, and Edward Dennis. Thematic review series: Proteomics. an integrated omics analysis of eicosanoid biology. *Journal of Lipid Research*, 50(6):1015–1038, 2009.

[10] Allan Butterfield and Barry Halliwell. Oxidative stress, dysfunctional glucose metabolism and alzheimer disease. *Nature Reviews Neuroscience*, 20(3):148–160, 2019.

[11] Huan Cai, Wei na Cong, Sunggoan Ji, Sarah Rothman, Stuart Maudsley, and Bronwen Martin. Metabolic dysfunction in alzheimer's disease and related neurodegenerative disorders. *Current Alzheimer research*, 9(1):5–17, 2012.

[12] Diogo Camacho, Alberto de la Fuente, and Pedro Mendes. The origin of correlations in metabolomics data. *Metabolomics*, 1(1), 2005.

[13] Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1):6, 2020.

[14] Davide Chicco, Niklas Tötsch, and Giuseppe Jurman. The matthews correlation coefficient (mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14(1):13, 2021.

[15] M.Azhar Chishti, Dun-Shen Yang, Christopher Janus, Amie L. Phinney, Patrick Horne, Jacqueline Pearson, Robert Strome, Noah Zuker, James Loukides, Janet French, Sherry Turner, Gianluca Lozza, Mariagrazia Grilli, Suzanne Kunicki, Céline Morissette, Julie Paquette, Francine Gervais, Catherine Bergeron, Paul E. Fraser, George A. Carlson, Peter St. George-Hyslop, and David Westaway. Early-onset amyloid deposition and cognitive deficits in transgenic mice expressing a double mutant form of amyloid precursor protein 695. *The Journal of Biological Chemistry*, 276(24):21562–21570, 2001.

[16] Lucía Chávez-Gutiérrez and Maria Szaruga. Mechanisms of neurodegeneration - insights from familial alzheimer's disease. *Seminars in Cell Developmental Biology*, 105:75–85, 2020.

[17] Llewelyn Davies, Beau Wolska, Caroline Hilbich, Gerd Multhaup, Ralph Martins, Simms, Konrad Beyreuther, and Colin Masters. A4 amyloid protein deposition and the diagnosis of alzheimer's disease: Prevalence in aged brains determined by immunocytochemistry compared with conventional neuropathologic techniques. *Neurology*, 38(11):1688, 1988.

[18] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240. Association for Computing Machinery, 2006.

[19] Suzanne de la Monte and Ming Tong. Brain metabolic dysfunction at the core of alzheimer's disease. *Biochemical Pharmacology*, 88(4):548–559, 2014.

[20] Paula de Matos, Rafael Alcántara, Adriano Dekker, Marcus Ennisa, Janna Hastings, Kenneth Haug, Inmaculada Spiteri, Steve Turner, and Christoph Steinbeck. Chemical entities of biological interest: an update. *Nucleic Acids Research*, 38:249–254, 2010.

[21] Suzana de Siqueira Santos, Daniel Yasumasa Takahashi, Asuka Nakata, and André Fujita. A comparative study of statistical methods used to identify dependencies between gene expression signals. 15(6):906–918, 2014.

[22] Lloyd Demetrius and Jane Driver. Alzheimer's as a metabolic disease. *Biogerontology*, 14(6):641–649, 2013.

[23] Edward Dennis, Raymond Deems, Richard Harkewicz, Oswald Quehenberger, Alex Brown, Stephen Milne, David Myers, Christopher Glass, Gary Hardiman, Donna Reichart, Alfred Merrill, Cameron Sullards, Elaine Wang, Robert Murphy, Christian Raetz, Teresa Garrett, Ziqiang Guan, Andrea Ryan, David Russell, Jeffrey McDonald, Bonne Thompson, Walter Shaw, Manish Sud, Yihua Zhao, Shakti Gupta, Mano Maurya, Eoin Fahy, and Shankar Subramaniam. A mouse macrophage lipidome. *The Journal of Biological Chemistry*, 285(51):39976–39985, 2010.

[24] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2):185–205, 2005.

[25] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004.

[26] Kieren Egan and Malcolm Macleod. Two decades testing interventions in transgenic mouse models of alzheimer's disease: designing and interpreting studies for clinical trial success. *Clinical Investigation*, 4:693–704, 2014.

[27] David Ellis, Warwick Dunn, Julian Griffin, William Allwood, and Royston Goodacre. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*, 8(9):1243–1266, 2007.

[28] Jeremiah Faith, Boris Hayete, Joshua Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James Collins, and Timothy Gardner. Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5:e8, 2007.

[29] Sarah Feldt, Paolo Bonifazi, and Rosa Cossart. Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends in Neurosciences*, 34(5):225–236, 2011.

[30] Cleusa Ferri, Martin Prince, Carol Brayne, Henry Brodaty, Laura Fratiglioni, Mary Ganguli, Kathleen Hall, Kazuo Hasegawa, Hugh Hendrie, Yueqin Huang, Anthony Jorm, Colin Mathers, Paulo Menezes, Elizabeth Rimmer, and Marcia Scazufca. Amyloid plaque core protein in alzheimer disease and down syndrome. *The Lancet*, 366(9503):2112–2117, 2005.

[31] Linton Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[32] George Glenner and Caine Wong. Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and Biophysical Research Communications*, 120(3):885–890, 1984.

[33] Michel Goedert and Maria Garzia Spillantini. A century of alzheimer's disease. *Science (New York, N.Y.)*, 314(5800):777–781, 2006.

[34] Marko Gosak, Rene Markovič, Jurij Dolenšek, Marjan Slak Rupnik, Marko Marhl, Andraž Stožer, and Matjaž Perc. Network science of biological systems at different scales: A review. *Physics of Life Reviews*, 24:118–135, 2018.

[35] Matthew Granger, Bettina Franko, Matthew Taylor, Claude Messier, Peter St George-Hyslop, and Steffany Bennett. A tgcrnd8 mouse model of alzheimer's disease exhibits sexual dimorphisms in behavioral indices of cognitive reserve. *Journal of Alzheimer's disease: JAD*, 51(3):757–773, 2016.

[36] Matthew W. Granger, Hui Liu, Caitlin F. Fowler, Alexandre P. Blanchard, Matthew W. Taylor, Samantha P. M. Sherman, Hongbin Xu, Weidong Le, and Steffany A. L. Bennett. Distinct disruptions in land's cycle remodeling of glycerophosphocholines in murine cortex mark symptomatic onset and progression in two alzheimer's disease mouse models. *Journal of Neurochemistry*, 149(4):499–517, 2019.

[37] Jan Grau, Ivo Grosse, and Jens Keilwagen. Prroc: computing and visualizing precision-recall and receiver operating characteristic curves in r. *Bioinformatics*, 31(15):2595–2597, 2015.

[38] Christine Grienberger, Nathalie L. Rochefort, Helmuth Adelsberger, Horst A. Henning, Daniel N. Hill, Julia Reichwald, Matthias Staufenbiel, and Arthur Konnerth. Staged decline of neuronal function in vivo in an animal model of alzheimer's disease. *Nature Communications*, 3(1):774, 2012.

[39] Karimollah Hajian-Tilaki. Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2):627–635, 2013.

[40] Chongomweru Halimu, Asem Kasem, and S. H. Shah Newaz. Empirical comparison of area under roc curve (auc) and mathew correlation coefficient (mcc) for evaluating machine learning algorithms on imbalanced datasets for binary classification. *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, pages 1–6, 2019.

[41] Xianlin Han. Lipidomics for studying metabolism. *Nature Reviews. Endocrinology*, 12(11):668–679, 2016.

[42] Violette Hanna and Ebtisam Abdel Aziz Hafez. Thematic review series: Proteomics. an integrated omics analysis of eicosanoid biology. *Journal of Advanced Research*, 11:23–32, 2018.

[43] Anne-Claire Haury, Fantine Mordelet, Paola Vera-Licona, and Jean-Philippe Vert. Tigress: Trustful inference of gene regulation using stability selection. *BMC Systems Biology*, 6(1):145, 2012.

[44] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi—a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[45] Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE*, 9(5), 2010.

[46] Trey Ideker and Nevan J Krogan. Differential network biology. *Molecular Systems Biology*, 8, 2012.

[47] Sanjeevan Jahagirdar and Edoardo Saccenti. On the use of correlation and mi as a measure of metabolite—metabolite association for network differential connectivity analysis. *Metabolites*, 10(4):171, 2020.

[48] Sanjeevan Jahagirdar, Maria Suarez-Diez, and Edoardo Saccenti. Simulation and reconstruction of metabolite-metabolite association networks using a metabolic dynamic model and correlation based algorithms. *Journal of Proteome Research*, 18(3):1099–1113, 2019.

[49] Natalia Janson. Non-linear dynamics of biological systems. *Contemporary Physics*, 53, 2012.

[50] Vinícius Carvalho Jardim, Suzana de Siqueira Santos, Andre Fujita, and Marcos Silveira Buckeridge. Bionetstat: A tool for biological networks differential analysis. *Frontiers in Genetics*, 10, 2019.

[51] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[52] Pablo Juanes-Velasco, Norma Galiciaa, Elisa Pin, Ricardo Jara-Acevedo, Javier Carabias-Sánchez, Rodrigo García-Valiente, Quentin Lecrevisse, Carlos Eduardo Pedreira, Rafael Gongora, Jose Manuel Sanchez-Santos, Héctor Lorenzo-Gil, Alicia Landeira-Viñuela, Halin Bareke, Alberto Orfao, Peter Nilsson, and Manuel Fuentes. Deciphering biomarkers for leptomeningeal metastasis in malignant hemopathies (lymphoma/leukemia) patients by comprehensive multi-pronged proteomics characterization of cerebrospinal fluid. *Cancers*, 14(2):449, 2022.

[53] Minoru Kanehisa and Susumu Goto. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1).

[54] Somang Kang, Yong ho Lee, and Jong Eun Lee. Metabolism-centric overview of the pathogenesis of alzheimer's disease. *Yonsei Medical Journal*, 58(3):479–488, 2017.

[55] M. Kendall. A new measure of rank correlation. 1938.

[56] Masashi Kitazawa, Rodrigo Medeiros, and Frank M. Laferla. Transgenic mouse models of alzheimer disease: developing a better model as a tool for therapeutic interventions. *Current Pharmaceutical Design*, 18(8):1131–1147, 2012.

[57] Mikaela Koutrouli, Evangelos Karatzas, David Paez-Espino, and Georgios A. Pavlopoulos. A guide to conquer the biological network era using graph theory. *Frontiers in Bioengineering and Biotechnology*, 8:34, 2020.

[58] Jan Krumsiek, Karsten Suhre, Thomas Illig, Jerzy Adamski, and Fabian J Theis. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Systems Biology*, 5:21, 2011.

[59] Bridget Kuehn. In alzheimer research, glucose metabolism moves to center stage. *JAMA*, 323(4):297–299, 2020.

[60] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[61] Rajeev Kumar and Abhaya Indrayan. Receiver operating characteristic (roc) curve for medical researchers. *Indian Pediatrics*, 48(4):277–287, 2011.

[62] Miyako Kusano, Atsushi Fukushima, Masanori Arita, Pär Jonsson, Thomas Moritz, Makoto Kobayashi, Naomi Hayashi, Takayuki Tohge, and Kazuki Saito. Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in arabidopsis thaliana. *BMC Systems Biology*, 1(1):53, 2007.

[63] Vincent Lacroix, Ludovic Cottret, Patricia Thébault, and Marie-France Sagot. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(4):594–617, 2008.

[64] C. A. Lane, J. Hardy, and J. M. Schott. Alzheimer's disease. *European Journal of Neurology*, 25(1):59–70, 2018.

[65] Cristina Licari, Leonardo Tenori, Betti Giusti, Elena Sticchi, Ada Kura, Rosina De Cario, Domenico Inzitari, Benedetta Piccardi, Mascia Nesi, Cristina Sarti, Francesco Arba, Vanessa Palumbo, Patrizia Nencini, Rossella Marcucci, Anna Maria Gori, Claudio Luchinata, and Edoardo Saccenti. Analysis of metabolite and lipid association networks reveals molecular mechanisms associated with 3-month mortality and poor functional outcomes in patients with acute ischemic stroke after thrombolytic treatment with recombinant tissue plasminogen activator. *Journal of Proteome Research*, 20(10):4758–4770, 2021.

[66] Chia-Chen Liu, Chia-Chan Liu, Takahisa Kanekiyo, Huaxi Xu, and Guojun Bu. Apolipoprotein e and alzheimer disease: risk, mechanisms and therapy. *Nature Reviews. Neurology*, 9(2):106–118, 2013.

[67] Chuang Liu, Yifang Ma, Jing Zhao, Ruth Nussinov, Yi-Cheng Zhang, Feixiong Cheng, and Zi-Ke Zhang. Computational network biology: Data, models, and applications. *Physics Reports*, 846:1–66, 2020.

[68] Gabriele Lohmann, Daniel S. Margulies, Annette Horstmann, Burkhard Pleger, Joeran Lepsien, Dirk Goldhahn, Haiko Schloegl, Michael Stumvoll, Arno Villringer, and Robert Turner. Eigenvector centrality mapping for analyzing connectivity patterns in fmri data of the human brain. *PLOS ONE*, 5(4), 2010.

[69] Hong-Wu Ma and An-Ping Zeng. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*, 19(11):1423–1430, 2003.

[70] Nivedhitha Mahendran, P. M. Durai Raj Vincent, Kathiravan Srinivasan, and Chuan-Yu Chang. Improving the classification of alzheimer's disease using hybrid gene selection pipeline and deep learning. *Frontiers in Genetics*, 12, 2021.

[71] Tom Maniatis and Robin Reed. An extensive network of coupling among gene expression machines. *Nature*, 416(6880):499–506, 2002.

[72] Daniel Marbach, James C. Costello, Robert Küffner, Nicole M. Vega, Robert J. Prill, Diogo M. Camacho, Kyle R. Allison, Manolis Kellis, James J. Collins,

and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, 2012.

[73] Adam Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(1):142, 2006.

[74] Caren Marzban. The roc curve and the area under it as performance measures. *Weather and Forecasting*, 19(6):1106–1114, 2004.

[75] Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 2002.

[76] C. L. Masters, G. Simms, N. A. Weinman, G. Multhaup, B. L. McDonald, and K. Beyreuther. Amyloid plaque core protein in alzheimer disease and down syndrome. *Proceedings of the National Academy of Sciences*, 82(12):4245–4249, 1985.

[77] Brian Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442–451, 1975.

[78] Aurélien Mazurie, Danail Bonchev, Benno Schwikowski, and Gregory A Buck. Evolution of metabolic network organization. *BMC Systems Biology*, 4:59, 2010.

[79] Peter Meikle, Gerard Wong, Christopher Barlow, and Bronwyn Kingwell. Lipidomics: potential role in risk prediction and therapeutic monitoring for diabetes and cardiovascular disease. *Pharmacology Therapeutics*, 143(1):12–23, 2014.

[80] Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

[81] Pedro Mendes, Stefan Hoops, Sven Sahle, Ralph Gauges, Joseph Dada, and Ursula Kummer. Computational modeling of biochemical networks using copasi. *Methods in Molecular Biology*, 500:17–59, 2009.

[82] Mario Mendez. Early-onset alzheimer disease and its variants. *Continuum (Minneapolis, Minn.)*, 25(1):34–51, 2019.

[83] Patrick Meyer, Kevin Kontos, Frederic Lafitte, and Gianluca Bontempi. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP journal on bioinformatics systems biology*, 2007.

[84] Leonor Michaelis, Maud Leonora Menten, Kenneth A. Johnson, and Roger S. Goody. The original michaelis constant: translation of the 1913 michaelis-menten paper. *Biochemistry*, 50(39):8264–8269, 2011.

[85] Lennart Mucke. Alzheimer's disease. *Nature*, 461(7266):895–897, 2009.

[86] Karl Pearson. Notes on the history of correlation. *Biometrika*, 13(1):25–45, 1920.

[87] Hanchuan Peng, Fuhui Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[88] Christian Raetz, Teresa Garrett, Michael Reynolds, Walter Shaw, Jeff Moore, Dale Smith, Anthony Ribeiro, Robert Murphy, Richard Ulevitch, Colleen Fearns, Donna Reichart, Christopher Glass, Chris Benner, Shankar Subramaniam, Richard Harkewicz, Rebecca Bowers-Gentry, Matthew Buczynski, Jennifer Cooper, Raymond Deems, and Edward Dennis. Kdo2-lipid a of escherichia coli, a defined endotoxin that activates macrophages via tlr-4. *Journal of Lipid Research*, 47(5):1097–1111, 2006.

[89] Antonio Rosato, Leonardo Tenori, Marta Cascante, Pedro Ramon De Atauri Carulla, Vitor A. P. Martins dos Santos, and Edoardo Saccenti. From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*, 14(4):37, 2018.

[90] Nadia Rosenthal and Steve Brown. The mouse ascending: perspectives for human-disease models. *Nature Cell Biology*, 9(9):993–999, 2007.

[91] Edoardo Saccenti, Maria Suarez-Diez, Claudio Luchinat, Claudio Santucci, and Leonardo Tenori. Probabilistic networks of blood metabolites in healthy sub-

jects as indicators of latent cardiovascular risk. *Journal of Proteome Research*, 14(2), 2014.

[92] Edoardo Saccenti, Maria Suarez-Diez, Claudio Luchinat, Claudio Santucci, and Leonardo Tenori. Probabilistic networks of blood metabolites in healthy subjects as indicators of latent cardiovascular risk. *Journal of Proteome Research*, 14(2):1101–1111, 2015.

[93] Michael M. Saint-Antoine and Abhyudai Singh. Network inference in systems biology: recent developments, challenges, and applications. *Current Opinion in Biotechnology*, 63:89–98, 2020.

[94] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), 2015.

[95] Ida Schomburg, Antje Chang, Sandra Placzek, Carola Söhngen, Michael Rother, Maren Lang, Cornelia Munaretto, Susanne Ulas, Michael Stelzer, Andreas Grote, Maurice Scheer, and Dietmar Schomburg. Brenda in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in brenda. *Nucleic Acids Research*, 41:764–772, 2031.

[96] Dennis Selkoe and John Hardy. The amyloid hypothesis of alzheimer's disease at 25 years. *EMBO molecular medicine*, 8(6):595–608, 2016.

[97] Diana W. Shineman, Guriqbal S. Basi, Jennifer L. Bizon, Carol A. Colton, Barry D. Greenberg, Beth A. Hollister, John Lincecum, Gabrielle G. Leblanc, Linda (Bobbi) H. Lee, Feng Luo, Dave Morgan, Iva Morse, Lorenzo M. Refolo, David R. Riddell, Kimberly Scearce-Levie, Patrick Sweeney, and Juha Yrjänheikki. Accelerating drug discovery for alzheimer's disease: best practices for preclinical animal studies. *Alzheimer's Research Therapy*, 3:28, 2011.

[98] Vladimir Shulaev. Metabolomics technology and bioinformatics. *Briefings in Bioinformatics*, 7(2):128–139, 2006.

[99] Lin Song, Peter Langfelder, and Steve Horvath. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics*, 13(1):328, 2012.

[100] Andrew Southam, John Easton, Grant Stentiford, Christian Ludwig, Theodoros Arvanitis, and Mark Viant. Metabolic changes in flatfish hepatic tumours revealed by nmr-based metabolomics and metabolic correlation networks. *Journal of Proteome Research*, 7(12):5277–5285, 2008.

[101] C. Spearman. 'general intelligence,' objectively determined and measured. *The American Journal of Psychology*, 15(2):201–293, 1904.

[102] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.

[103] Gustavo Stolovitzky, Don Monroe, and Andrea Califano. Dialogue on reverse-engineering assessment and methods. *Annals of the New York Academy of Sciences*, 1115(1):1–22, 2007.

[104] Maria Suarez-Diez and Edoardo Saccenti. Effects of sample size and dimensionality on the performance of four algorithms for inference of association networks in metabonomics. *Journal of Proteome Research*, 14:5119–5130, 2015.

[105] Neil Swainston, Kieran Smallbone, Hooman Hefzi, Paul D. Dobson, Judy Brewer, Michael Hanscho, Daniel C. Zielinski, Kok Siong Ang, Natalie J. Gardiner, Jahir M. Gutierrez, Sarantos Kyriakopoulos, Meiyappan Lakshmanan, Shangzhong Li, Joanne K. Liu, Veronica S. Martínez, Camila A. Orellana, Lake-Ee Quek, Alex Thomas, Juergen Zanghellini, Nicole Borth, Dong-Yup Lee, Lars K. Nielsen, Douglas B. Kell, Nathan E. Lewis, and Pedro Mendes. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics: Official Journal of the Metabolomic Society*, 12:109, 2016.

[106] Jedrzej Szymanskik, Szymon Jozefczuk, Zoran Nikoloski, Joachim Selbig, Victoria Nikiforova, Gareth Catchpole, and Lothar Willmitzer. Stability of metabolic correlations under changing environmental conditions in escherichia coli – a systems approach. *PLOS ONE*, 4(10):e7441, 2009.

[107] Jon Toledo, Matthias Arnold, Gabi Kastenmüuller, Rui Chang, Rebecca Baillie, Xianlin Han, Madhav Thambisetty, Jessica Tenenbaum, Karsten Suhre, Will Thompson, Lisa St. John-Williams, Siamak MahmoudianDehkordi, Daniel Rotroff, John Jack, Alison Motsinger-Reif, Shannon Risacher, Colette Blach,

Joseph Lucas, Tyler Massaro, Gregory Louie, Hongjie Zhu, Guido Dallmann, Kristaps Klavins, Therese Koal, Sungeun Kim, Kwangsik Nho, Li Shen, Ramon Casanova, Sudhir Varma, Cristina Legido-Quigley, Arthur Moseley, Kuixi Zhu, Marc Henrion, Sven van der Lee, Amy Harms, Ayse Demirkan, Thomas Hankemeier, Cornelia van Duijn, John Trojanowski, Leslie Shaw, Andrew Saykin, Michael Weiner, Murali Doraiswamy, and Rima Kaddurah-Daouk. Metabolic network failures in alzheimer's disease: A biochemical road map. *Alzheimer's dementia : the journal of the Alzheimer's Association*, 13(9):965–984, 2017.

[108] Eugenia Trushina and Michelle M. Mielke. Recent advances in the application of metabolomics to alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842:1232–1239, 2014.

[109] Peter Uetz, Loic Giot, Gerard Cagney, Traci A. Mansfield, Richard S. Judson, James R. Knight, Daniel Lockshon, Vaibhav Narayan, Maithreyan Srinivasan, Pascale Pochart, Alia Qureshi-Emili, and Ying Li Brian Godwin Diana Conover Theodore Kalbfleisch Govindan Vijayadamodar Meijia Yang Mark Johnston Stanley Fields Jonathan M. Rothberg. A comprehensive analysis of protein–protein interactions in saccharomyces cerevisiae. *Nature*, 403(6770):623–627, 2000.

[110] Hanna Vesterinen, Emily Sena, Charles ffrench Constant, Anna Williams, Siddharthan Chandran, and Malcolm Macleod. Improving the translational hit of experimental treatments in multiple sclerosis. *Multiple Sclerosis*, 16:1044–1055, 2010.

[111] Alessia Vignoli, Leonardo Tenori, Betti Giusti, Serafina Valente, Nazario Carrabba, Daniela Balzi, Alessandro Barchielli, Niccolò Marchionni, Gian Franco Gensini, Rossella Marcucci, Anna Maria Gori, Claudio Luchinat, and Edoardo Saccenti. Differential network analysis reveals metabolic determinants associated with mortality in acute myocardial infarction patients and suggests potential mechanisms underlying different clinical scores used to predict death. *Journal of Proteome Research*, 19(2):949–961, 2020.

[112] Alessia Vignoli, Leonardo Tenori, Claudio Luchinat, and Edoardo Saccenti. Differential network analysis reveals molecular determinants associated with blood pressure and heart rate in healthy subjects. *Journal of Proteome Research*, 20(1):1040–1051, 2021.

[113] Bei Wang, Lujin Wu, Jing Chen, Lingli Dong, Chen Chen, Zheng Wen, Jiong Hu, Ingrid Fleming, and Dao Wen Wang. Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. *Signal Transduction and Targeted Therapy*, 6(1):1–30, 2021.

[114] Rand Wilcox. *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, 2012.

[115] Jordan Maximillian Wilkins and Eugenia Trushina. Application of metabolomics in alzheimer's disease. *Frontiers in Neurology*, 8:719, 2018.

[116] Claude Wischik, Michal Novák, Hans Christian Thøgersen, Patricia Edwards, Michael Runswick, Ross Jakes, Walker, César Milstein, Martin Roth, and Aaron Klug. Isolation of a fragment of tau derived from the core of the paired helical filament of alzheimer disease. *Proceedings of the National Academy of Sciences*, 85(12):4506–4510, 1988.

[117] D. Wishart. Current progress in computational metabolomics. *Briefings Bioinform.*, 2007.

[118] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, Olga Krebs, Meik Bittkowski, Elina Wetsch, Isabel Rojas, and Wolfgang Müller. Sabio-rk—database for biochemical reaction kinetics. *Nucleic Acids Research*, 40:790–796, 2012.

[119] Ivan Rodrigo Wolf, Rafael Plana SimõesRafael Plana Simões, and Guilherme Targino Valente. Three topological features of regulatory networks control life-essential and specialized subsystems. *Scientific Reports*, 11(1):24209, 2021.

[120] Naomi Wraya, Jian Yang, Michael Goddard, and Peter Visscher. The genetic interpretation of area under the roc curve in genomic profiling. *PLOS Genetics*, 6(2), 2010.

[121] Kui Yang and Xianlin Han. Lipidomics: Techniques, applications, and outcomes related to biomedical sciences. *Trends in biochemical sciences*, 41(11):954–969, 2016.

[122] Chun-Hou Zheng, Lin Yuan, Wen Sha, and Zhan-Li Sun. Gene differential coexpression analysis based on biweight correlation and maximum clique. *BMC Bioinformatics*, 15, 2015.

[123] Qiuming Zhu. On the performance of matthews correlation coefficient (mcc) for imbalanced dataset. *Pattern Recognition Letters*, 136:71–80, 2020.