applied sciences

MDPI

*Article*

# Mutual Information Based on Multiple Level Discretization Network Inference from Time Series Gene Expression Profiles

**Cao-Tuan Anh and Yung-Keun Kwon ***

Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan 44610, Republic of Korea; tacao@mta.edu.vn
* Correspondence: kwonyk@ulsan.ac.kr

**Abstract:** Discovering a genetic regulatory network (GRN) from time series gene expression data plays an essential role in the field of biomedical research. In its development, many methods have been proposed for inferring GRNs. Although most of them are effective, they have limitations in terms of network size and the number of regulatory genes due to high computational cost. Thus, it is necessary to develop an efficient method that can operate with large networks and provide reliable results within an acceptable run time. In this study, we propose a new method using mutual information based on multi-level discretization network inference (MIDNI) from time series gene expression profiles. The proposed method discretizes time series gene expression data to minimize information loss and computational consumption through K-means clustering. We do not fix the number of clusters, instead varying it depending on the distribution of gene expression values. We compared MIDNI with three well-known inference methods through extensive simulations on both artificial and real gene expression datasets. Our results illustrate that MIDNI significantly outperforms the alternatives in terms of dynamic accuracy. The proposed method represents an efficient and scalable tool for inferring GRNs from time series gene expression data.

**Keywords:** gene regulatory network; gene network inference; multiple-level discretization; time-series gene expression

check for updates

**Citation:** Anh, C.-T.; Kwon, Y.-K. Mutual Information Based on Multiple Level Discretization Network Inference from Time Series Gene Expression Profiles. *Appl. Sci.* **2023**, *13*, 11902. https://doi.org/10.3390/app132111902

Academic Editor: Keun Ho Ryu

Received: 24 September 2023
Revised: 21 October 2023
Accepted: 26 October 2023
Published: 31 October 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

A gene regulatory network (GRN) is a collection of molecular regulators that interact with each other and other substances in the cell to govern the gene expression levels of mRNA and proteins. Uncovering the underlying interactions genetic from a time series gene expression dataset is a network inference or reverse engineering problem, and many methods have been proposed to address it. Specifically, in this problem the goal involves finding a set of regulatory genes for each target gene as well as inferring the regulatory rules between genes.

Data-driven methods are a class of GRN reconstruction methods that estimate gene dependencies directly from the data. Within this class, the correlation score is widely used to associate a pair of vector-valued measurements. Weighted gene co-expression network analysis (WGCNA) [1] is a well known correlation score-based method that shows consistently reliable performance. However, the correlation coefficients fail to capture more complex statistical dependencies, e.g., non-linear ones, between expression patterns. To resolve this limitation, the mutual information (MI) has been used as an alternative. MI is an efficient information-theoretic score that has been frequently used in the determination of regulatory relations. The relevance network is the simplest model based on this measure; it computes MI between all pairs of genes and infers the presence of a regulatory interaction when the MI is larger than a given threshold. The context likelihood of relatedness (CLR) algorithm [2] is an extension of the relevance network approach used to replace the probability distributions of all MI scores with empirical

distributions. CLR applies an adaptive background correction step to eliminate false correlations and indirect influences. The algorithm for the reconstruction of accurate cellular networks (ARACNE) [3] is another information-theoretic algorithm for reverse engineering transcriptional networks from microarray data. ARACNE defines an edge as an irreducible statistical dependency between gene expression profiles that cannot be explained as an artifact of other statistical dependencies in the network. Finally, minimum redundancy networks (MRNET) [4] use an effective information-theoretic technique for feature selection based on a maximum relevance/minimum redundancy criterion. Despite promising results, computational complexity remains a major limitation. Another limitation is that most of these methods are only applicable to smaller networks.

In order to reduce computational complexity and speed up running time, many methods to transform real-valued gene expression data into binary data have been widely used. These binarized expression data are then used as an input for inferring systems. The best fit [5] is an example method that considers the consistency as well as best-fit extension problems in the context of inferring networks from discretized data. Another example of using binary data [6] is to infer a Boolean network (BN) model of a GRN from limited transcriptomic or proteomic time series data. This method is able to illustrate the inference of a BN from limited time series data with constraints on connectivity that explain the observed state transitions. The mutual information-based Boolean network inference (MIBNI) method [7] uses Boolean data discretized from time series expression data as an input for inferring regulatory networks. MIBNI first identifies a set of initial regulatory genes using mutual information-based feature selection, then improves the dynamics prediction accuracy by iteratively swapping a pair of genes between sets of the selected regulatory genes and the other genes. A genetic algorithm-based Boolean network inference (GABNI) method [8] is an upgraded approach from MIBNI. GABNI applies a genetic algorithm (GA) to search an optimal set of regulatory genes in a wider solution space when MIBNI fails to find an optimal solution in a small-scale inference problem. Although GABNI shows good performance for large-scale networks, it employs a limited representation model of regulatory functions. In this regard, a novel genetic algorithm combined with a neural network for the Boolean network inference was proposed in [9], where a neural network was used to represent the regulatory function instead of the incomplete Boolean truth table used in the GABNI. In brief, although the inferred results were reliable with acceptable accuracy, Boolean network methods often show poor performance due to the simplicity (information loss) of the Boolean representation of the data.

To extend the limited representation of the Boolean network for the network reconstruction problem, we propose a novel algorithm called mutual information based on multiple-level discretization network inference from gene expression profiles (MIDNI). The algorithm first discretizes gene expression data with multiple levels (two or three) according to the distribution of the expression values of a gene. Through analysis based on the validity index [10,11] over massive data, we observe that two-level discretization performs best on certain genes. However, three-level discretization is shown to be effective for a considerably larger number of genes. On the other hand, using a value of $k$ larger than three increases system complexity and computational cost. Subsequently, the discretized expression data are fed to the mutual information–based feature selection (MISF) method [7]. We used MIFS to estimate the mutual information between a target gene and a set of candidate regulatory genes to reduce the computational cost. The loss of information increases as the number of regulatory genes considered in MIFS increases, which can eventually cause a significant decrease in inference accuracy. To overcome this problem with MIFS, we used a SWAP subroutine, which is a greedy algorithm wherein a gene in the set of regulatory genes selected by MIFS is iteratively swapped with another gene in the set of unselected genes. MIFS and SWAP were presented in the MIBNI algorithm [7]. However, MIFS and SWAP in MIBNI were implemented to process only Boolean expression data. Thus, we carry out appropriate modifications to these subroutines. We validate the performance

of our method on both artificial and real gene expression datasets in comparison with three well-known methods, namely, new dynamic Bayesian network (DBN) [12], maximal information coefficient with conditional relative average entropy and time series mutual information (MICRAT) [13], and the aforementioned MIBNI [7]. Our experimental result show that our method outperforms all three in terms of dynamic accuracy. Furthermore, MIDNI is able to infer a network against noisy datasets. This indicates that MIDNI is a notable method for inferring regulatory networks from gene expression datasets.

## 2. Materials and Methods

### 2.1. Related Works

In this subsection, we introduce three relevant methods: DBN, MICRAT, and MIBNI. These are later intensively compared with our proposed method in the experimental results section.

The new DBN method [12] is a method for reconstructing genetic networks from time series gene expression data, and is based on previous DBN family methods [14–19]. The main improvement in the newest version is that it limits the potential regulators to those genes with either earlier or simultaneous expression changes (up- or downregulation) in relation to their target genes. Consequently, the search space for potential regulators is reduced. Furthermore, for a given regulator and target gene, the time between the initial change in the expression is tuned to estimate the transcriptional time lag between these two genes. The MICRAT method similarly infers GRNs from time series gene expression data [13]. The Maximal Information Coefficient (MIC) [20] is used to measure the dependence between two genes in both functional and non-functional associations. This approach reconstructs an undirected graph to represent the underlying relationships between genes as well as the direction for inferring regulators and their targets. Finally, MIBNI is an MI-based Boolean network inference method [7]. For each target gene in the network, this method first identifies a set of candidate regulators using a mutual information-based feature selection. Subsequently, pairs of genes consisting of sets made up of selected regulatory genes and other genes are iteratively swapped to improve the dynamic prediction accuracy.

### 2.2. Discretized Network Model

In this study, we employed a gene-wise discretized network model to investigate the complex dynamics of gene regulatory networks. A discretized network is represented by a directed graph $G(V, A)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of nodes, $A = \{(v_i, v_j)\} \subseteq V \times V$ is a set of interactions, and the state value of gene $v$ at time $t$, $v(t)$ is represented by $l$ discrete values $\{0, 1, 2, \ldots, l-1\}$. We note that this is called a Boolean network if $l = 2$ for all genes in $V$. Consider a target node $v \in V$ regulated by $k$ genes $u_1, u_2, \ldots, u_k (u_i \in V)$. Let $E$ and $E_i$ be the sets of discretized expression values of genes $v$ and $u_i$, respectively. The value $v(t+1)$ is updated by a discrete function $f : E_1 \times E_2 \times \cdots \times E_k \to E$ of the values of $k$ regulatory genes $u_1, u_2, \ldots, u_k$ at time $t$. Hence, the update scheme of $v$ can be described by the following formula:

$$v(t+1) = f(u_1(t), u_2(t), \ldots, u_k(t)) \quad (1)$$

We note that the update time lag used in this study is one and that the number of all possible functions with respect to $f$ is $l^{\prod_i l_i}$, where $l$ and $l_i$ are the cardinalities of $E$ and $E_i$, respectively.

### 2.3. The Discretization Network Inference Problem

The network inference problem is the problem of inferring both a set of interactions and a set of update functions from time series gene expression data. The inference performance can be evaluated by comparing the trajectory generated by the inferred network and the observed time series gene expression. Let $v^{'}(t)$ be the predicted value of gene $v$ at time $t$ in the inferred discretization network. We define the consistency of the gene-wise

dynamics $C(v, v^{'})$ as the similarity between the discretization trajectories of the observed gene expression $v(t)$ and the estimated gene expression $v^{'}(t)$, as follows:

$$C(v, v^{'}) = \frac{\sum_{t=p+1}^{T} I(v(t) = v^{'}(t))}{T - p}, \tag{2}$$

where $T$ is the total number of time steps, $p$ is the time lag, and $I(\cdot)$ is an indicator function that returns 1 if the condition is true and otherwise returns 0. In addition, the comparison starts at $t = 2$, as $p$ is set to 1 in this study. Finally, we define the dynamic accuracy of an inferred network as the average of the gene-wise dynamics over all genes, as follows:

$$DynamicsAccuracy = \frac{\sum_{i=1}^{N} C(v_i, v_i^{'})}{N}, \tag{3}$$

where $N$ is the number of genes.

### 2.4. Structure Performance Metrics

When the structure of a gold standard or correct network is known, it is possible to further evaluate the inference performance with respect to the network structure. To this end, we use three measures: precision, recall, and structural accuracy. Precision is the ratio of correctly inferred connections over the total number of predictions:

$$Precision = \frac{TP}{TP + FP}, \tag{4}$$

where $TP$ (true positive) and $FP$ (false positive) denote the numbers of correctly and incorrectly predicted connections, respectively. Recall is the ratio of true predicted connections over the total number of actual connections:

$$Recall = \frac{TP}{TP + FN}, \tag{5}$$

where $FN$ (false negative) means the number of non-inferred connections in $G(V, A)$. Structural accuracy is the ratio of correct predictions out of all predictions:

$$StructuralAccuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{6}$$

where $TN$ (true negative) is the number of correct negative predictions.

## 3. Our Proposed Method

In this paper, we propose a novel method called MIDNI for a multiple-level discretization network inference from gene expression data. Figure 1 demonstrates the overall structure of MIDNI. A time series gene expression dataset is used as input, and is converted into a gene expression dataset using the K-means discretization method [21–23]. All expression values of each gene are divided into two or three clusters depending on the distribution of gene expression. When the maximum discretization level is two, the values of genes are marked by 0 (low) and 1 (high). On the other hand, when it is three they are marked by 0, 1, and 2, denoting low, typical, and high expression levels, respectively. For each target gene $v$ of which the entropy value is non-zero, the MIFS subroutine is used to select $k$ genes in $V$ that have the most informative variables with gene $v$. Then, the SWAP subroutine is applied to improve the consistency of the gene-wise dynamic by swapping the same number of variables between $S$ and $W \setminus S$. This procedure is repeated by increasing the number $k$ until an optimal set $S$ is found (the gene-wise dynamic consistency of gene $v$ equals 1) or $k$ reaches a user-defined parameter $K$ that indicates the maximum number of regulatory genes for gene $v$ to be inferred.

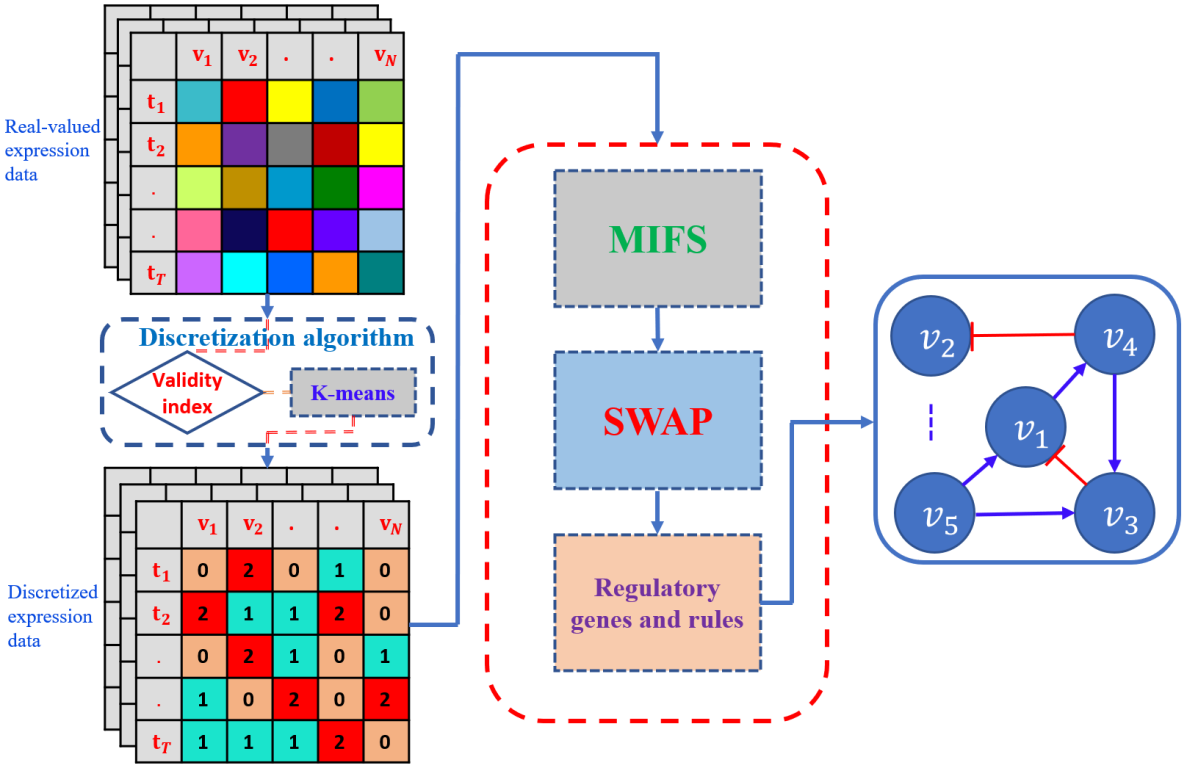


**Figure 1.** Framework of the MIDNI algorithm. A time series gene expression dataset (the columns denote genes as $v_1, v_2, \ldots, v_N$ and the rows denote time steps as $t_1, t_2, \ldots, t_N$) is used as input. The validity index is used to determine the most suitable number of clusters for each gene. Subsequently, the discretized expression dataset generated by the K-means method is used as input to an inference algorithm consisting of MIFS and SWAP subroutines. Finally, the regulatory network is reconstructed based on the obtained regulatory genes and rules.

### *3.1. Discretization*

Boolean discretization, which uses only two states (0 for off and 1 for on) is very popular, as it can speed up the computational process and save on running time. However, the conversion from real-valued expression data to Boolean data often reduces the accuracy of the inference system due to information loss and representative simplicity. In this regard, it is often appropriate to convert the real-valued expression of certain genes into a higher-level discretized expression values. Hence, MIDNI employs a hybrid approach for data discretization considering both two- and three-level discretizations. To determine the optimal level of discretization, we used the validity index [10,11,24]. The validity index is defined as

$$Validity = \frac{Intra}{Inter}, \tag{7}$$

where

$$Intra = \frac{1}{T}\sum_{i=1}^{C}\sum_{x\in C_i} \| x - z_i \|^2 \text{ and}$$

$$Inter = \min_{i\neq j\in\{1,2,\cdots,k\}} \| z_i - z_j \|^2 .$$

In the above equations, $T$ is the number of time steps, $C$ is the number of clusters, and $z_i$ denotes the center of the $i$th cluster $C_i$. The optimal number of clusters is chosen as $C$ to minimize the validity index. The expression values of a gene are classified into two or three classes by K-means clustering, then converted to the discrete values corresponding to the cluster.

### *3.2. MIFS and SWAP Subroutines*

Our proposed MIDNI algorithm employs two subroutines, MIFS and SWAP, similar to MIBNI [7]. However, in the MIBNI algorithm MIFS and SWAP only work with the binary expression. Thus, we modified them to adapt both two-level and three-level discretized expressions, as shown in Algorithms 1 and 2. For each target gene $v$ of the network, the MIFS subroutine returns an initial set of candidate regulatory genes $S_v \subseteq V$ by evaluating an approximated multivariate mutual information. Subsequently, the SWAP subroutine is used to increase the dynamic accuracy by swapping between selected candidate regulatory genes $S_v$ and the set of unselected genes $V \setminus S_v$. This swapping process is repeated until there is no improvement in the consistency of the gene-wise dynamics. Finally, MIDNI predicts a set of regulatory genes for each target gene of the input network. Based on this result, the inference method is applied to reconstruct the original network.

**Algorithm 1** $MIFS(v_0, W, k)$ subroutine, where $v_0$ is the target variable, $W = \{w_1, w_2, ..., w_M\}$ is the set of variables, and $k$ is the desired number of input variables

**Require:** $k \geq 1$, $|W| \geq 1$
  $S \leftarrow$ Initialize to empty
  $v \leftarrow argmax_{w \in W} I(v_0; w)$
  $S \leftarrow S \cup v$
  $W \leftarrow W \setminus v$
  **while** $|S| \leq k$ **do**
    $v \leftarrow argmax_{w \in W}(I(v_0; w) - \Sigma_{s \in S} I(w; s))$
    $S \leftarrow S \cup v$
    $W \leftarrow W \setminus v$
  **end while**
  **return** $S$

**Algorithm 2** $SWAP(v_0, S, W)$ subroutine, where $v_0$ is the target variable $S = s_1, s_2, ..., s_k$ is the set of selected variables such that $I(v_0; s_i) \geq I(v_0; s_j)$ if $i < j$ for all $s_i, s_j \in S$, and $W = \{w_1, w_2, ..., w_M\}$ is the set of unselected variables such that $I(v_0; w) \geq I(v_0; w_j)$ if $i < j$ for all $w_i, w_j \in W$

  $v_0^{'} \leftarrow search_update_rule(v_0, S)$
  $E_{MAX} \leftarrow E(v_0, v_0^{'})$
  $i = 1$
  **while** $i \leq M$ **do**
    $j = 1$
    **while** $j \leq K$ **do**
      $v_0^{'} \leftarrow search_update_rule(v_0, S \cup \{w_j\} \setminus \{s_j\})$
      **if** $(E(v_0, v_0^{'}) > E_{MAX})$ **then**
        $S \leftarrow S \cup \{w_i\}$
        $S \leftarrow S \setminus \{s_j\}$
        $W \leftarrow W \cup \{s_j\}$
        $W \leftarrow W \setminus \{w_j\}$
        $E_{MAX} \leftarrow E(v_0, v_0^{'})$
      **end if**
    **end while**
  **end while**
  **return** $(S, E_{MAX})$

## 4. Results

To validate the performance of MIDNI, we tested it on two gene expression datasets from an artificial dataset and an *Escherichia coli* gene regulatory network generated by GeneNetWeaver (GNW) [25].

### *4.1. Case Study 1: Artificial Dataset*

We generated twenty groups of artificial datasets with network sizes = 10, 20, . . . , 190, and 200. In each group, twenty different networks were randomly generated, for a total of 400 tested networks. The state of each gene (node) was randomly initialized among 0, 1, and 2 and updated over 29 time steps using an update function selected uniformly and randomly from a set of update functions. The set of update functions contained a huge number of possible update functions for each target gene regulated by $k$ other genes, and no discretization method was required for this dataset. See Table 1 for details about the settings of the generated dataset system. MIDNI inferred a network structure and a set of predicted regulatory rules for all genes. Finally, we analyzed the performance of MIDNI in terms of both structural and dynamic accuracy.

**Table 1.** The setup parameters used to generate the dataset.

| Parameters | Artificial Dataset | *Escherichia coli* Dataset [25] |
|---|---|---|
| Size of networks ($N$) | 10, 20, . . . , 190, 200 | 10, 50, 100, 200 |
| Noise rate | without noise | 0, 5, 10% |
| Time lag | 1 | 1 |
| Number of time points ($T$) | 30 | 30 |
| Include at least regulators | $0.4 \times N$ | $0.4 \times N$ |

The experimental results of MIDNI are shown in Figure 2. As shown in the figure, the dynamics accuracy, precision, and recall decrease as the size of the network increases. Contrary to these three metrics, the structural accuracy gradually rises as the size of the network increases. This is because the number of true positives declines as the network size increases, whereas the number of true negatives increases significantly. Furthermore, we examined the effect of the number of incoming links on performance (Figure 3). The number of incoming links specifies the difficulty of the inference problem. When there is only one incoming link for a target gene (the easiest case), our method always predicts correctly. However, the performance metrics gradually decrease as the number of incoming links increases.

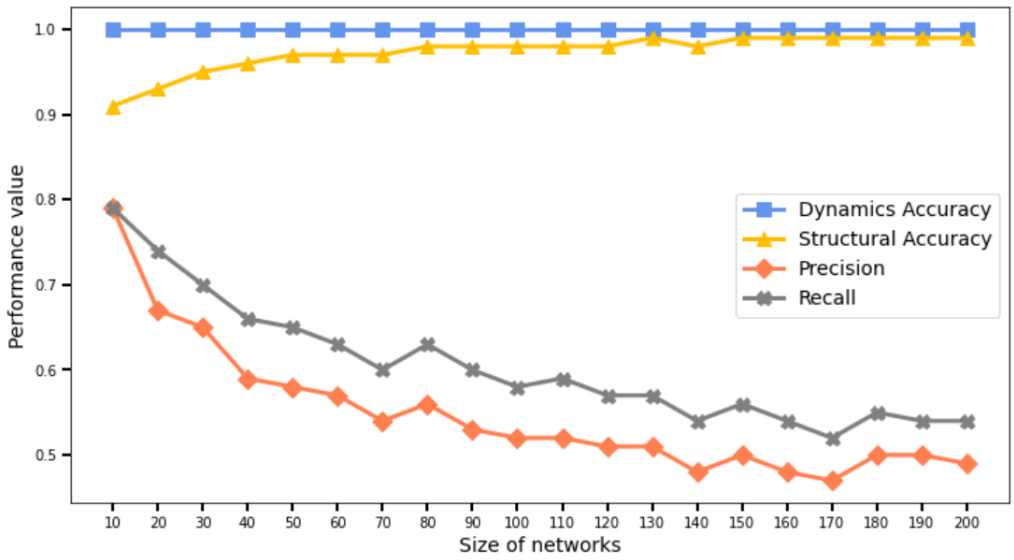


**Figure 2.** Performance of MIDNI on the artificial dataset. The y-axis denotes the performance values for precision, recall, dynamic accuracy, and structural accuracy (see Sections 2.3 and 2.4), while the x-axis represents the network size (the number of genes), which varies from 10 to 200.

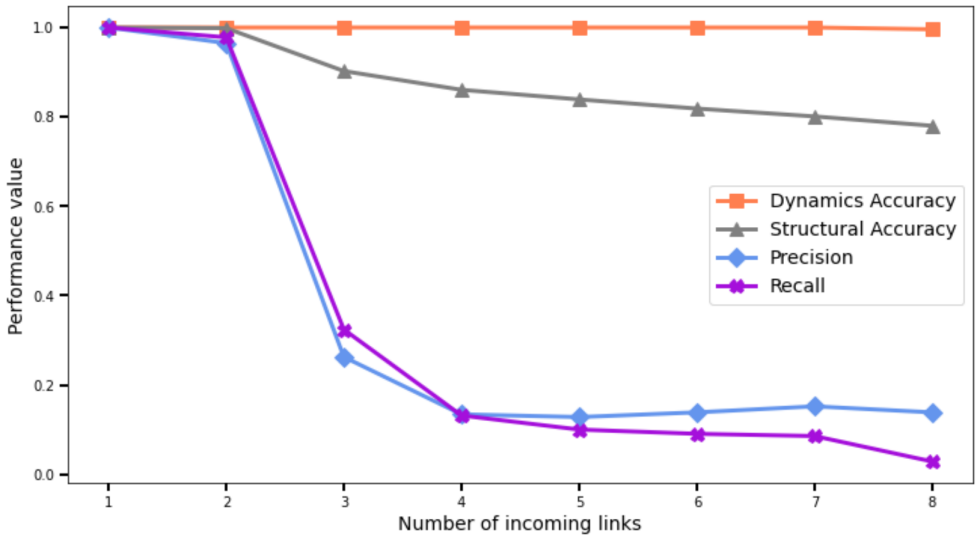


**Figure 3.** The effect of the number of incoming links on the inference result on the artificial dataset. The y-axis denotes the performance values and the x-axis represents the number of incoming links, specifying the difficulty of the inference problem.

### *4.2. Case Study 2: Escherichia Coli Dataset*

Next, we tested our method on a time series gene expression dataset generated by the GNW tool. The settings used for GNW are shown in Table 1. We used four groups of datasets with sizes of 10, 50, 100, and 200. Each group contained three subgroups divided according to noise rates, and each subgroup had twenty different networks, for a total of 240 tested networks. After the dataset was generated, the expression values were clustered using the K-means method, with $k$ equaling two or three depending on the distribution of the values. Based on the discretization dataset, MIDNI inferred the network structure and a set of regulatory rules for all genes. We analyzed its performance in terms of both the structural and dynamic accuracy.

**Structural accuracy analysis.** We compared the structural accuracy (see Equation (6)) of MIDNI with the DBN, MICRAT, and MIBNI methods on the *Escherichia coli* dataset (Figure 4). As shown in the figure, the structural accuracy of MIDNI was lower than that of the other methods, particularly for the networks with the smallest size ($|V| = 10$). In fact, the structural accuracy was affected by the number of positive predictions (=TP + FP in 2.4) of the inference method. Specifically, the structural accuracy is likely to become higher as the number of positive predictions decreases, as a GRN is a considerably sparse graph. Figure 5 shows the comparison results for the number of positive predictions among the inference methods. As shown in the figure, MIDNI predicted the existence of interactions significantly more often than the other methods. Therefore, it should be noted that its low structural accuracy represents a trade-off for the numerous positive predictions.

**Dynamic accuracy analysis.** It is well known that networks with different structures can produce the same underlying dynamics. Hence, it is essential to verify the network inference performance in terms of the dynamic accuracy. In particular, the dynamic accuracy is not affected by the number of positive predictions, unlike the structural accuracy. We compared the dynamic accuracy (see Equation (3)) of MIDNI, DBN, MICRAT, and MIBNI on the *Escherichia coli* dataset (Figure 6). Figure 6a–c shows the results on the dataset with noise rates of 0%, 5%, and 10%, respectively. As shown in Figure 6a, MIDNI demonstrated considerably higher dynamic accuracy than the other methods regardless of the network size. This implies that MIDNI shows robust performance in terms of the dynamic accuracy against the network size. In addition, the dynamic accuracy of MIDNI was consistently higher than the other methods in the case of noisier datasets, as shown in Figure 6b,c. This implies that MIDNI is robust against noise in gene expression values.

Finally, we examined the proportions of genes classified by two- or three-level discretization in each network (Figure 7). As shown in the figure, the proportions of the two- and three-level discretized genes were comparable to each other except for the networks with size 10. This result explains why the multi-level discretization should be considered to improve dynamic accuracy compared to the more commonly used binarization approach.

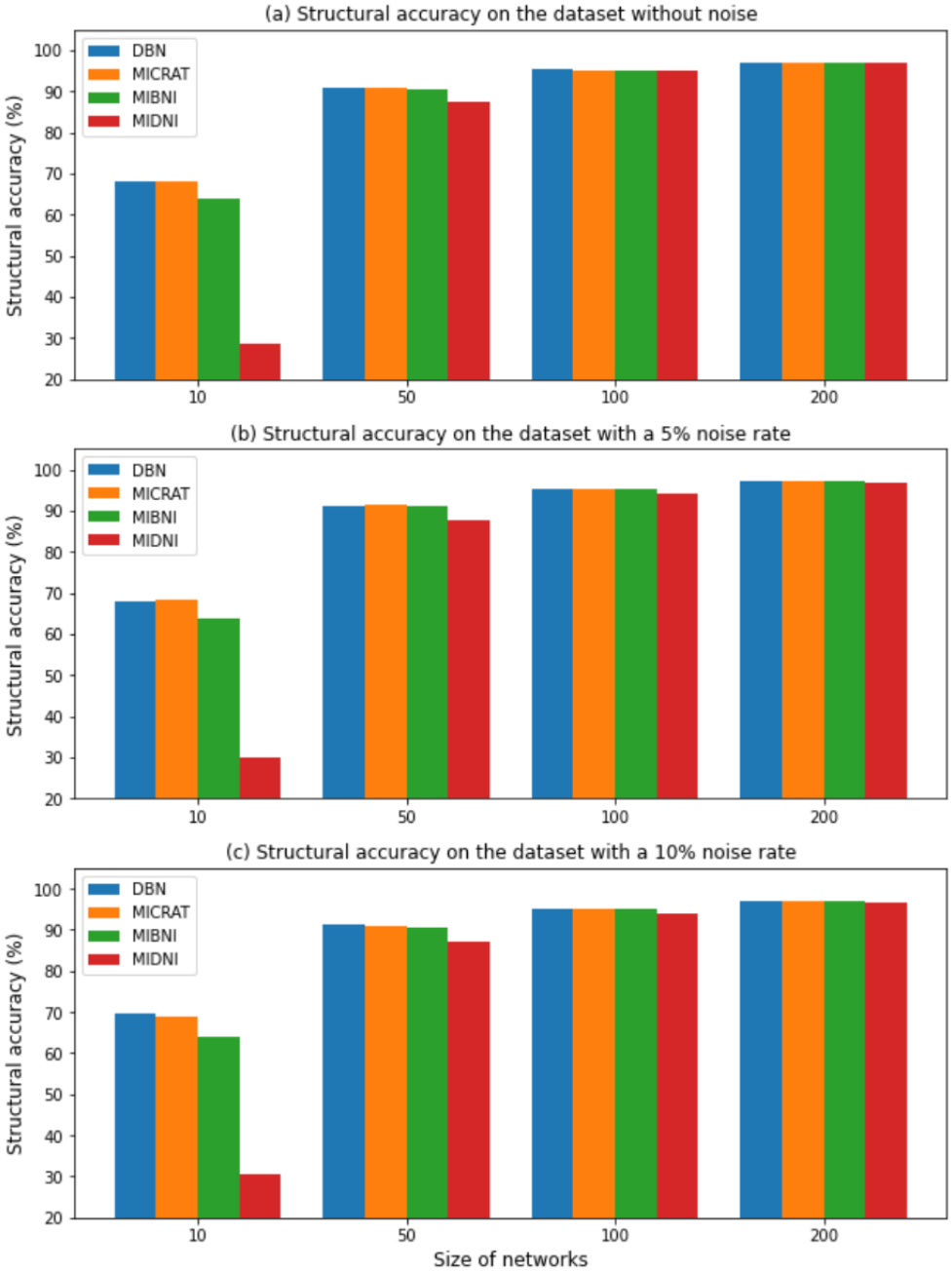


**Figure 4.** Structural accuracy comparison on the *E. coli* dataset.

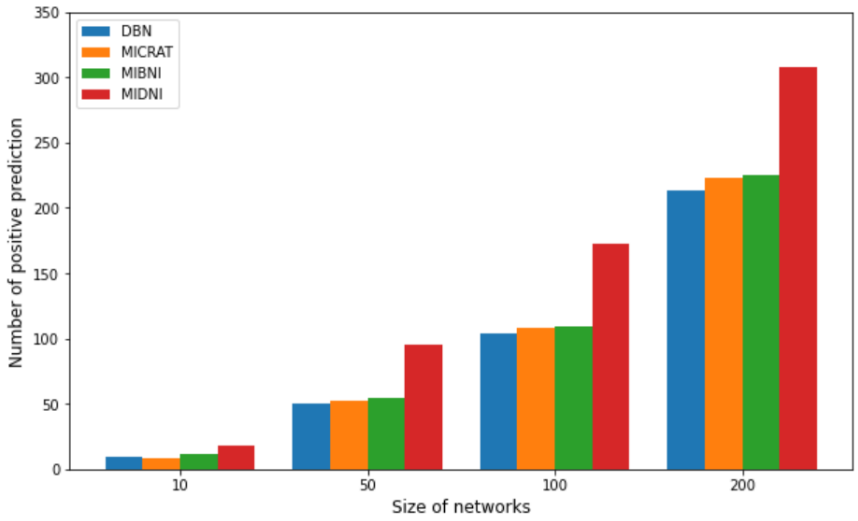


**Figure 5.** Comparison of the number of positive predictions on the *E. coli* dataset. The x-axis denotes the size of the networks, while the y-axis denotes the average number of positive predictions for each target gene.

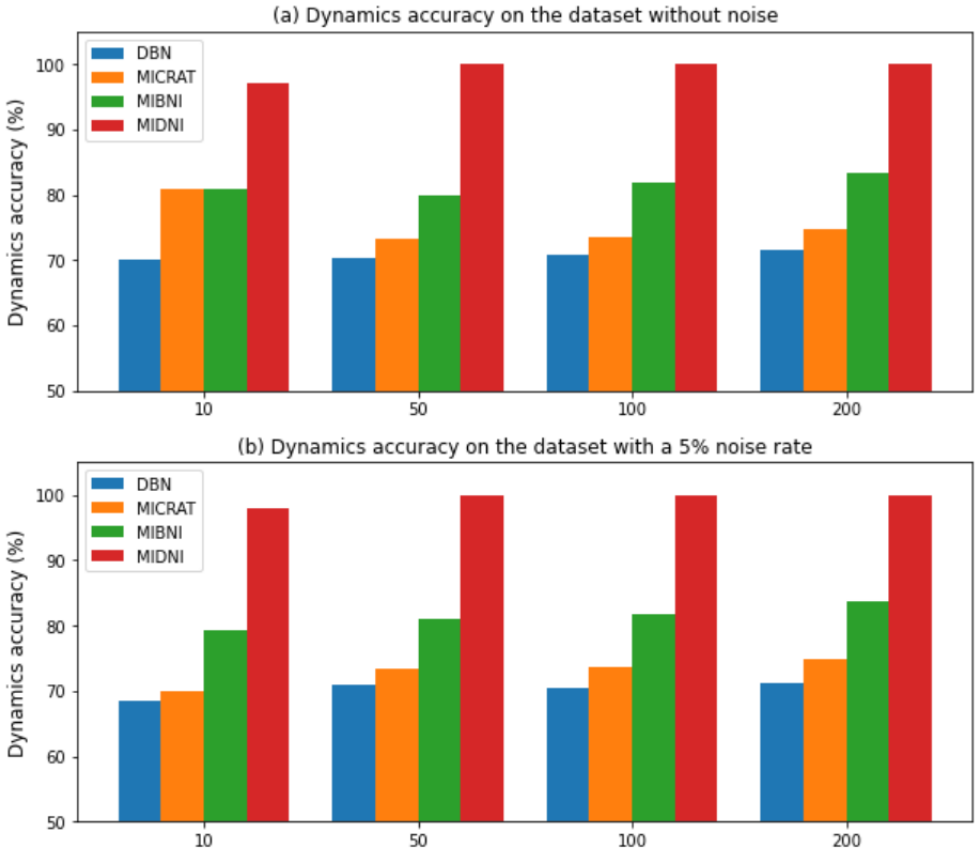


**Figure 6.** *Cont.*

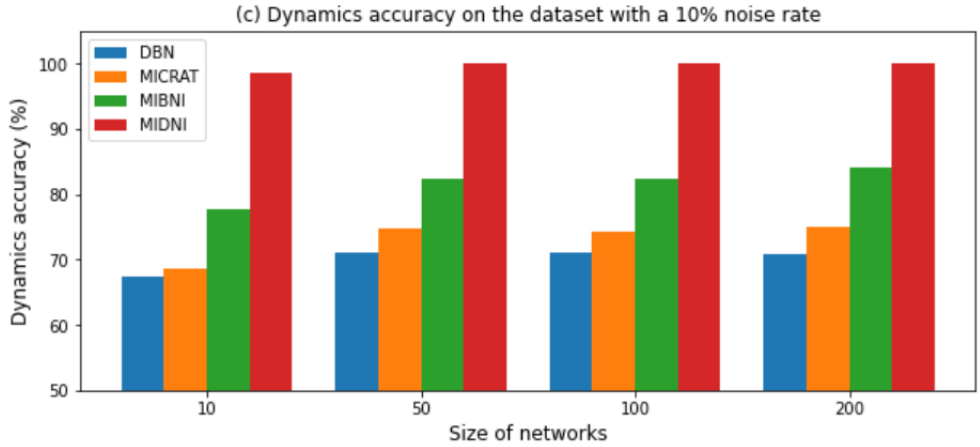


**Figure 6.** Dynamic accuracy comparison on the *E. coli* dataset.

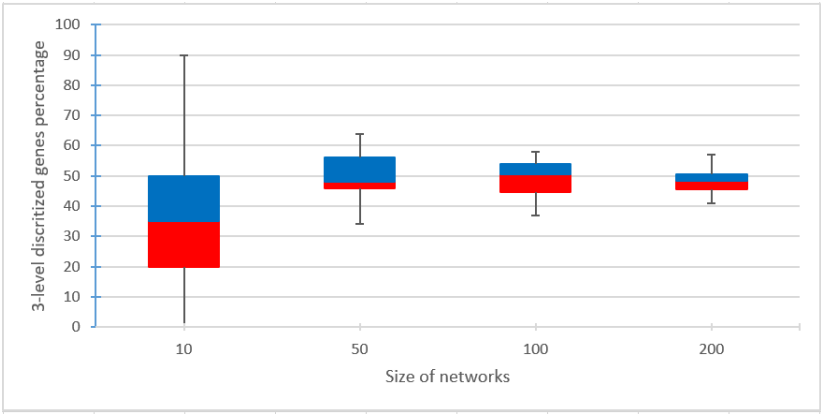


**Figure 7.** Distribution of three-level discretized genes on the *E. coli* dataset.

**Running time analysis.** To compare the running time of MIDNI to the other methods, we examined the average running time over a total of 180 gene expression datasets on a PC with an AMD Ryzen 5 3400 G 3.7 GHz CPU and 16 GB of RAM (Figure 8). As shown in the figure, we observed the running time of MIDNI to be higher than those of BDN and MICRAT, whereas it was comparable to that of MIBNI. While our method is not faster than the others, the increasing slope of the running time against the network size is almost same among all the methods. In addition, the running time is only slightly influenced by the noise rate. Considering that the dynamic accuracy of MIDNI overwhelmed the other methods, it can be concluded that our method improves performance by sacrificing a reasonable amount of running time.

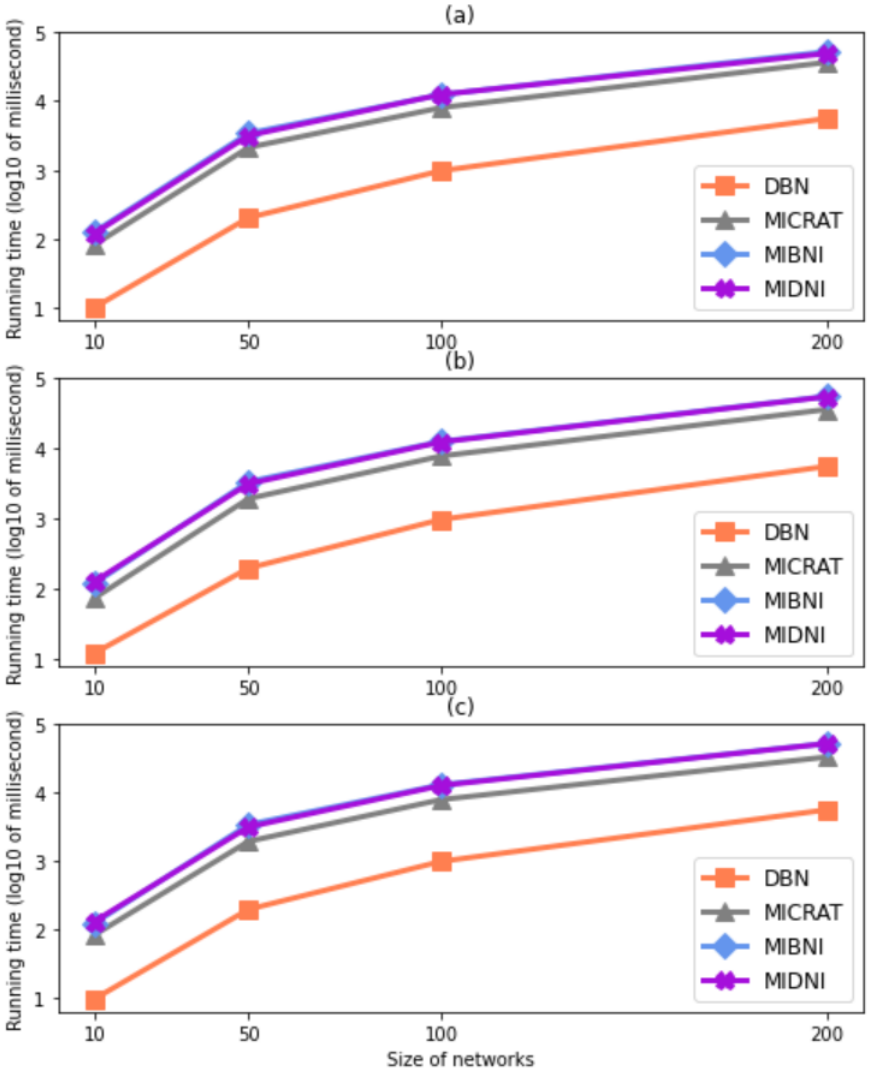


**Figure 8.** Running time comparison on the *E. coli* dataset: (**a**–**c**) show the results on the dataset with noise rates of 0%, 5%, and 10% respectively.

## 5. Discussion and Conclusions

In this study, we have proposed a new network inference method for use with time series gene expression data, which we call MIDNI. Although many previous methods have been proposed, most deal with real-valued or Boolean expression data. Therefore, they are limited in terms of the applicable network size or experience performance degradation due to reliance on simple representations. Our method implements a multi-level discretization network inference model, and outperforms the compared methods (DBN, MICRAT, and MIBNI). Specifically, MIDNI showed the best results in terms of dynamics accuracy. However, MIDNI has limitations that need to be improved. First, its results in terms of structural accuracy are low, as the current approach relies on the mutual information to select a set of candidate regulatory genes. For certain genes which interact with others via an intermediate gene, the mutual information can be larger than that of genes which directly interact, potentially leading to erroneous positive predictions. Second, MIDNI uses a correlation coefficient to determine the interaction direction of genes. This approach is relatively simple, and there is a room for improvement. Finally, as the MIFS and SWAP subroutines used in MIDNI are greedy algorithms, the run time could be further improved by developing more efficient search algorithms.

**Author Contributions:** Formal analysis, C.-T.A.; Writing—original draft, C.-T.A.; Writing—review & editing, Y.-K.K.; Supervision, Y.-K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the 2023 Research Fund of the University of Ulsan.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** https://github.com/tacaomta/MIDNISourceCode.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zhang, B.; Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef]
2. Faith, J.J.; Hayete, B.; Thaden, J.T.; Mogno, I.; Wierzbowski, J.; Cottarel, G.; Kasif, S.; Collins, J.J.; Gardner, T.S. Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol.* **2007**, *5*, e8. [CrossRef]
3. Zoppoli, P.; Morganella, S.; Ceccarelli, M. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinform.* **2010**, *11*, 154. [CrossRef]
4. Meyer, P.E.; Kontos, K.; Lafitte, F.; Bontempi, G. Information-Theoretic Inference of Large Transcriptional Regulatory Networks. *Eurasip J. Bioinform. Syst. Biol.* **2007**, *2007*, 1–9. [CrossRef]
5. Lähdesmäki, H.; Shmulevich, I.; Yli-Harja, O. On Learning Gene Regulatory Networks Under the Boolean Network Model. *Mach. Learn.* **2003**, *52*, 147–167. [CrossRef]
6. Haider, S.; Pal, R. Boolean network inference from time series data incorporating prior biological knowledge. *BMC Genom.* **2012**, *13*, S9. [CrossRef]
7. Barman, S.; Kwon, Y.-K. A novel mutual information-based Boolean network inference method from time-series gene expression data. *PLoS ONE* **2017**, *12*, e0171097. [CrossRef]
8. Barman, S.; Kwon, Y.K. A Boolean network inference from time-series gene expression data using a genetic algorithm. *Bioinformatics* **2018**, *34*, i927–i933. [CrossRef]
9. Barman, S.; Kwon, Y.K. A neuro-evolution approach to infer a Boolean network from time-series gene expressions. *Bioinformatics* **2020**, *36* (Suppl. S2), i762–i769. [CrossRef]
10. Shen, J.; Chang, S.I.; Lee, E.S.; Deng, Y.; Brown, S.J. Determination of cluster number in clustering microarray data. *Appl. Math. Comput.* **2005**, *169*, 1172–1185. [CrossRef]
11. Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. Vazirgiannis, Clustering algorithms and validity measures. In Proceedings of the Thirteenth International Conference on Scientific and Statistical Database Management, Fairfax, VA, USA, 18–20 July 2001.
12. Zou, M.; Conzen, S.D. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **2004**, *21*, 71–79. [CrossRef]
13. Yang, B.; Xu, Y.; Maxwell, A.; Koh, W.; Gong, P.; Zhang, C. MICRAT: A novel algorithm for inferring gene regulatory networks using time series gene expression data. *BMC Syst. Biol.* **2018**, *12*, 115. [CrossRef]
14. de Luis Balaguer, M.A.; Sozzani, R. Inferring Gene Regulatory Networks in the Arabidopsis Root Using a Dynamic Bayesian Network Approach. *Methods Mol. Biol.* **2017**, *1629*, 331–348.
15. Yu, J.; Smith, V.A.; Wang, P.P.; Hartemink, A.J.; Jarvis, E.D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **2004**, *20*, 3594–3603. [CrossRef]
16. Imoto, S.; Goto, T.; Miyano, S. Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression. *Pac. Symp. Biocomput.* **2002**, *7*, 175–186.
17. Kim, S.Y.; Imoto, S.; Miyano, S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Briefings Bioinform.* **2003**, *4*, 228–235. [CrossRef]
18. Murphy, K.; Mian, S. *Modeling Gene Expression Data Using Dynamic Bayesian Networks*; Technical Report, Computer Science Division; University of California: Berkeley, CA, USA, 1999.
19. Perrin, B.-E.; Ralaivola, L.; Mazurie, A.; Bottani, S.; Mallet, J.; D'alché–Buc, F. Gene networks inference using dynamic Bayesian networks. *Bioinformatics* **2003**, *19*, ii138–ii148. [CrossRef]
20. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [CrossRef]
21. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Davis Davis, CA, USA, 21 June–18 July 1965 and 27 December–7 January 1966.
22. Bholowalia, P.; Kumar, A. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *Int. J. Comput. Appl.* **2014**, *105*, 17–24.

23. Kodinariya, T.M.; Makwana, P.R. Review on determining number of Cluster in K-Means Clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2013**, *1*, 90–95.
24. Ray, S.; Turi, R.H. Determination of number of clusters in K-means clustering and application in colour image segmentation. In Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDTÕ99), New Delhi, India, 27–29 December 1999; pp. 27–29.
25. Schaffter, T.; Marbach, D.; Floreano, D. GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* **2011**, *27*, 2263–2270. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.