Building Undirected Influence Ontologies Using Pairwise Similarity Functions

Tamlin Love
School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg
South Africa
tamlinlollislove@gmail.com

Ritesh Ajoodha
School of Computer Science and Applied Mathematics
The University of the Witwatersrand, Johannesburg
South Africa
ritesh.ajoodha@wits.ac.za

***Abstract*—The recovery of influence ontology structures is a useful tool within knowledge discovery, allowing for an easy and intuitive method of graphically representing the influences between concepts or variables within a system. The focus of this research is to develop a method by which undirected influence structures, here in the form of undirected Bayesian network skeletons, can be recovered from observations by means of some pairwise similarity function, either a statistical measure of correlation or some problem-specific measure.**

In this research, we present two algorithms to construct undirected influence structures from observations. The first makes use of a threshold value to filter out relations denoting weak influence, and the second constructs a maximum weighted spanning tree over the complete set of relations. In addition, we present a modification to the minimum graph edit distance (GED) [1], which we refer to as the modified scaled GED, in order to evaluate the performance of these algorithms in reconstructing known structures. We perform a number of experiments in reconstructing known Bayesian network structures, including a real-world medical network [2]. Our analysis shows that these algorithms outperform a random reconstruction (modified scaled GED ≈ 0.5), and can regularly achieve modified scaled GED scores better than 0.3 in sparse cases and 0.45 in dense cases.

We argue that, while these methods cannot replace traditional Bayesian network structure-learning techniques, they are useful as computationally cheap data exploration tools and in knowledge discovery over structures which cannot be modelled as Bayesian networks.

***Index Terms*—ontology, structure learning, knowledge discovery, minimum graph edit distance, graphical model**

I. Introduction

A common way of representing knowledge within a domain is through an ontology: a set of concepts in the domain and the relations between them [3]. A common form of ontology is the influence ontology, whose relations encode causal influence between the concepts or variables in the ontology. For example, an influence ontology may be employed to represent the influence relations between diseases and symptoms in a medical setting (see Figure 2 in Section IV-B).

The ability to recover the underlying structure of an influence ontology from observations is a useful tool in knowledge discovery. This task has been achieved to various degrees of success in the existing literature. Many methods exist to recover a structure alongside a conditional probability distribution (CPD), at the cost of limiting the structure to a directed acyclic model (see Section II-A) [4] [5] [6].

If instead we discard both direction of influence and the CPD, we can instead recover an undirected structure, as in [7] [8] [9]. These methods, however, make other assumptions, either restricting the model structure or assuming continuous, rather than categorical data. The purpose of this research, therefore, is to develop a computationally cheap method by which generalised undirected influence structures can be reconstructed from observations, either continuous or categorical, using pairwise similarity functions.

To that end, this research makes the following contributions. Firstly, we present two algorithms to reconstruct an undirected influence structure from observations under varying density assumptions. Both methods can make use of any pairwise similarity measure, either a statistical measure, as in this research, or a problem-specific similarity measure such as a distance metric or semantic similarity. Secondly, we present a modification to the graph edit distance (GED) for comparing undirected graphs over shared variable spaces, scaled by the number of possible edges, here termed the modified scaled GED, as a means of evaluating our recovered structure against a *ground truth* model.

The structure of this paper is as follows. In Section II, we provide backgrounds of key concepts related to our research. Section III contains the contributions of this research, namely our two proposed algorithms in Section III-A and our proposed evaluation score in Section III-B. In Section IV, we outline a set of experiments to test the efficacy of our algorithms. The results of these experiments are presented and discussed in Section V.

II. Background and Related Work

A. Modelling Ontologies

As ontologies consist of domain concepts or variables and the relations between them, a natural approach to modelling them is to use some form of graphical model, whose vertices represent the concepts or variables and whose edges represent the relations.

Bayesian networks are directed acyclic graphs (DAGs) whose vertices represent random variables and whose edges represent influence of one variable on another, coupled with a conditional probability distribution (see Figure 1) [10]. This

978-1-7281-4162-6/20$31.00 ©2020 IEEE

structure can also be thought of as a representation of the conditional independencies between the random variables [11]. Indeed, it is through the exploitation of these independency assumptions that a Bayesian network can more compactly represent a joint distribution.

When modelling an influence ontology, a Bayesian network presents two limitations. First of all, the direction assigned to each edge may not necessarily reflect the causal flow in the original system, and may instead result in a model which is I-equivalent to the *ground truth* model [13]. Secondly, limiting the structure to an acyclic model eliminates valid ontology structures (for example, a system in which two variables influence each other over time).

Given these restrictions, we may instead wish to generalise our graphical model to an undirected structure, at the cost of losing the CPDs and thus the ability to perform inference. Indeed, for the purposes of knowledge discovery, an undirected model can still provide useful information about the structure of an influence ontology. Nevertheless, Bayesian networks are still useful in this research, as they provide a method of sampling observations from a known *ground truth* structure.

B. Undirected Structure Learning

Several techniques have been used previously to learn the structure of undirected models. For example, one approach is to calculate the mutual information between each variable pair, sort these values and then construct edges in descending order provided no path already exists between two edges (to ensure no unnecessary edges are added) [7]. This is essentially Kruskal's algorithm [14], adapted to find the maximum weighted spanning tree of the graph, where each edge is weighted using the mutual information.

Another approach to continuous data is to assume it follows a Gaussian distribution with covariance Σ, which implies that if $\Sigma^{-1}_{ij} = 0$, then variables X_i and X_j are conditionally independent [8]. LASSO is then used to estimate the presence of non-zero elements in Σ^{-1}. An extension of this work is to use a coordinate descent method to estimate Σ^{-1} [9].

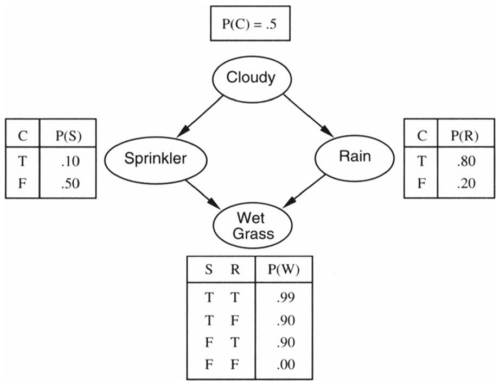


Fig. 1: A famous example of a Bayesian network, showing how a complete representation of any random variable X requires considering only those variables who are parents of X in the graphical representation [12].

C. Model Evaluation

When reconstructing a model, it is important to measure how well one's learned model recovers the *ground truth* structure. As the focus of this research is in learning the undirected structure of such models, any evaluation metric must rely solely on its structural features.

There are a few existing metrics for scoring the similarity between graphs. One of the most intuitive metrics is the minimum graph edit distance (GED) [15]. For two graphs, $\mathcal{G}_0$ and $\mathcal{G}_1$, the minimum graph edit distance, $GED(\mathcal{G}_0, \mathcal{G}_1)$, is defined as the minimum number of operations required to transform $\mathcal{G}_0$ into $\mathcal{G}_1$. For an unweighted and undirected graph where the set of operations consists of edge deletion, edge insertion, vertex deletion and vertex insertion, this is formulated as

$$GED(\mathcal{G}_0, \mathcal{G}_1) = \sum_{e \in E_0 - E_1} c_{ed} + \sum_{e \in E_1 - E_0} c_{ei} + \sum_{v \in V_0 - V_1} c_{vd} + \sum_{v \in V_1 - V_0} c_{vi} \quad (1)$$

where E_i and V_i are the sets of edges and vertices respectively in graph $\mathcal{G}_i$, and c_{ed}, c_{ei}, c_{vd} and c_{vi} are the costs of edge deletion, edge insertion, vertex deletion and vertex insertion respectively [1]. Assuming non-negative costs and noting that $GED(\mathcal{G}_i, \mathcal{G}_j) \geq 0$, we have that values of $GED(\mathcal{G}_i, \mathcal{G}_j)$ close to 0 indicate similar graphs, whereas high values of $GED(\mathcal{G}_i, \mathcal{G}_j)$ indicate very dissimilar graphs.

III. Research Methodology

In this research, we present two algorithms to reconstruct the undirected skeleton of an influence ontology over a set of N variables, $\mathcal{X}$, given a set of M observations $\mathcal{D} = \{\boldsymbol{o}[1], ..., \boldsymbol{o}[M]\}$ and a similarity function, $similarity$. We also present a modified version of the GED score presented in equation 1 to be used in our experiments in Section IV.

A. Algorithms

Our first algorithm relies on a threshold parameter $t \in [0, 1]$. In this simple approach, we first construct a similarity matrix $S_{ij} = similarity(X_i, X_j)$ and then keep any edges e_{ij} where $|S_{ij}| > t$. Here we take the absolute value to account for similarity functions whose range includes negative numbers. In pseudocode, the algorithm is as follows.

Algorithm 1 Build Influence Ontology - Threshold Approach

```
procedure BUILD_GRAPH(𝒳, 𝒟, similarity, t)
    𝒢 ← (V = 𝒳, E = ∅)
    N ← Dim(𝒢)
    S_ij ← similarity(𝒟, i, j)
    for i ∈ [0, ..., N − 1] do
        for j ∈ [0, ..., N − 1] do
            if i ≠ j and |S_ij| ≥ t then
                addEdge(𝒢, i, j)
    return 𝒢
```

An important consideration is the choice of threshold value. Clearly, a low t would favour a denser graph, while a high

t would favour a sparser graph. Some prior insight into the expected density of the *ground truth* model will therefore assist in the choice of t.

One potential limitation with this approach is that the similarity function merely measures similarity and does not encode conditional independence. Therefore similarities between vertices without a common edge but with a common ancestor may be mistaken for two vertices with influence between them. In these cases, it may be desirable to recover a sparse structure, such as a tree. To do this, we employ a similar methodology to [7], here using Kruskal's algorithm [14] to find the maximum weighted spanning tree over the complete graph weighted $e_{ij} = |S_{ij}|$, as in the following pseudocode.

Algorithm 2 Build Influence Ontology - MWST Approach

procedure BUILD_GRAPH($\mathcal{X}$, $\mathcal{D}$, $similarity$)
  $\mathcal{G} \leftarrow (V = \mathcal{X}, E = \emptyset)$
  $N \leftarrow Dim(\mathcal{G})$
  $S_{ij} \leftarrow similarity(\mathcal{D}, i, j)$
  for $i \in [0, ..., N-1]$ **do**
    for $j \in [0, ..., N-1]$ **do**
      if $i \neq j$ **then**
        $addEdge(\mathcal{G}, i, j)$ with $weight = |S_{ij}|$
  $\mathcal{T} \leftarrow Kruskal\_Get\_MWST(\mathcal{G})$
  return $\mathcal{T}$

Here there is no parameter, as the choice of density is implicit in the decision to recover a spanning tree. Obviously, limiting the reconstructed structure to a tree will preclude other potentially valid structures (for example, the structure in Figure 1). The spanning property may also add weak edges between nodes that, in the *ground truth* structure, belong to disconnected sub-graphs. However, the method has other strengths, as the undirected structure can be transformed into a directed structure by directing all edges away from a given vertex. The choice of root vertex is not obvious, but may be assisted by domain knowledge.

B. Modified Scaled Graph Edit Distance

To modify the GED score presented in equation 1, we begin by setting all costs to 1, and assuming $V_0 = V_1$. Assuming a symmetric adjacency matrix representation of $\mathcal{G}_0$ and ,$\mathcal{G}_1$ (with 0s in the main diagonal), equation 1 reduces to

$$GED(\mathcal{G}_0, \mathcal{G}_1) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} |E_{ij}^{(0)} - E_{ij}^{(1)}|, \qquad (2)$$

where $E_{ij}^{(k)}$ is the adjacency matrix of $\mathcal{G}_k$. As with equation 1, a score of 0 in this case represents a perfect reconstruction. As the number of edges in an undirected N-node graph is bounded above by $\frac{N(N-1)}{2}$ [16], the worst possible edit distance is $\frac{N(N-1)}{2}$. Thus, by dividing equation 2 by $\frac{N(N-1)}{2}$, we arrive at a modified scaled graph edit distance, a measure where 0 indicates the best possible reconstruction and 1 indicates the worst.

$$GED(\mathcal{G}_0, \mathcal{G}_1) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |E_{ij}^{(0)} - E_{ij}^{(1)}|, \qquad (3)$$

IV. Experiments

To test the efficacy of the algorithms presented in Section III, we detail an experiment in recovering the undirected *ground truth* structures of several Bayesian Networks of varying densities. In addition to this experiment, we also apply algorithms 1 and 2 to data sampled from a real Bayesian network constructed by domain experts.

A. Bayesian Network Reconstruction

While our algorithms can be applied to any graphical structure, Bayesian networks (whose structures are directed acyclic graphs) are chosen for the experiment in this research, as observations can be sampled from them, allowing for comparison against a known *ground truth* model.

To that end, we began by randomly generating 200 Bayesian networks. To do this, we firstly made use of the fact that a graph is acyclic if and only if it has a topological ordering [17], which is true if and only if the adjacency matrix can be transformed into a strongly upper triangular matrix [18], or equivalently, a strongly lower triangular matrix. To generate a random lower triangular adjacency matrix, we defined a density factor, ρ, and constructed each edge in the lower triangle with uniform probability $P(e_{ij}) = \rho$. Thus a low ρ is likely to produce a sparse graph, while a high ρ is likely to produce a dense graph.

For every $N \in \{4, 8, 12, 16, 20\}$ (where N denotes the number of vertices in the graph) and each $\rho \in \{0.2, 0.4, 0.6, 0.8\}$, we generated 10 graphs, thus resulting in 200 graphs across a range of variable space sizes and densities. These adjacency matrices were then used to construct continuous Bayesian networks with Gaussian CPDs and 20,000 observations were sampled from each network using forward sampling.

In our experiment, we used algorithms 1 and 2 to obtain structures for each graph (varying the threshold parameter t for algorithm 1). In this research, we made use of the Pearson correlation coefficient [19] and Spearman's rank correlation coefficient [20] as the similarity functions used to recover influence between the variables. The modified scaled GED score (equation 3) was then used to evaluate the graph against the symmetric version of the adjacency matrix used to generate the observations. The average of the 10 graphs that share N and ρ values was then computed and plotted against a varying t.

As a baseline for both experiments, we used a random graph reconstruction which randomly constructed edges with uniform probability $P(e_{ij}) = 0.5$.

B. CHILD Network Application

To test our methodology on a real application, we make use of the CHILD Network (Figure 2), a Bayesian network constructed by [21] and adapted by [2], used to diagnose congenital heart disease in infants suffering from "blue baby syndrome".

The network consists of 20 discrete nodes with 25 edges. 10,000 observations have been sampled from the network. As the variables are categorical, we employ the bias-corrected

Cramèr's V as our similarity measure [22]. Varying t for algorithm 1, we compute the modified scaled GED for both algorithms.

V. Results and Discussion

A. Bayesian Network Reconstruction

We begin by examining the modified GED Score against the threshold parameter t for the most sparse graphs ($\rho = 0.2$) for varying numbers of variables N (see Figures 3 and 4). The error bars in each of the plots in this research denote one standard deviation from the mean. For ease of viewing, the error bars for the random reconstruction have been omitted. Example reconstructions for each method can be found in Figure 5.

Our first observation is that both similarity functions produce almost identical results for the threshold approach (algorithm 1) and identical results for the MWST approach (algorithm 2). This trend continues throughout the results in this research, and points to a linear correlation between variable pairs.

Another observation is that for all values of $t \neq 0$, the threshold approach outperforms the random approach by a significant margin (as high as 0.3 GED for the 20-node case). The threshold approach curves all follow a "u-shape", with the dip of the curve shifting towards higher t values as N increases. The MWST approach also outperforms the random approach, with a higher margin as N increases. For larger values of N, the MWST approach achieves the best results, with a modified scaled GED score of 0.1242 for $N = 20$.

We now examine the most dense graphs ($\rho = 0.8$), varying t and N as before (see Figures 6 and 7). In these results, we see somewhat of a reversal of some of the previous trends. For all values of N, there are low values of t that cause the threshold approach to perform the best, with a modified scaled GED score of around 0.2. As with the sparse case, even the best scores increase as N grows larger. The MWST approach performs more poorly at this density, performing worse than the random reconstruction.

In summary, we observe the following. For sparse graphs, the MWST approach produces the best reconstructions of the *ground truth* models. This is probably because sparse graphs are likely to, at the very least, resemble tree structures. The threshold approach outperforms the random approach, with low t performing better for low N and high t performing better for high N.

For dense graphs, the threshold approach produces the best reconstructions of the *ground truth* models. This is because the approach is capable of capturing the dense structure and is not hindered by the limitation discussed in section III-A. Low t values tend to produce better reconstructions, as they allow for denser graphs. In contrast, the MWST approach performs poorly for dense graphs, as tree structures are, by their nature, sparse.

Overall, the algorithms perform best when reconstructing sparse structures. This is because sparse Bayesian networks encode fewer conditional dependencies, and thus variable pairs that do not have a direct edge are less likely to be strongly correlated.

B. CHILD Network Application

For this application, we begin by executing algorithms 1 and 2, varying t in the case of algorithm 1, and plotting the modified scaled GED against t, as in Figure 8.

Our results here are consistent with the results for a very sparse graph (e.g. Figure 4). The MWST approach produces the best reconstruction, with a score of 0.042105 (8 edits). The threshold approach yields a best score of 0.07368 (14 edits) for $t = 0.4715$. Figure 9 demonstrates these reconstructions.

In the best cases, both approaches tend to construct very sparse structures over the variables, with the threshold approach producing many disconnected components. The MWST approach produces a very close reconstruction, preserving many of the correct edges (apart from a reversal in the

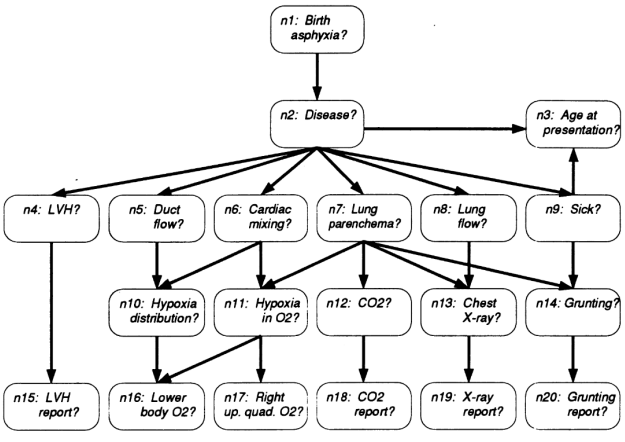


Fig. 2: The CHILD Network, a real-world Bayesian Network, used to diagnose congenital heart disease in infants suffering from "blue baby syndrome" [2]

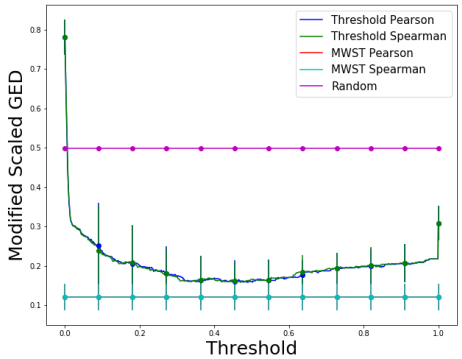


Fig. 3: Modified Scaled GED Score vs Threshold, $\rho = 0.2$, $N = 12$

CO2 and *CO2 Report* branch), and overall appears to capture the 6 distinct branches in the *ground truth* structure.

VI. Conclusion

The usefulness of our research depends largely on the requirements of the problem they are used to solve. We have shown empirically that algorithms 1 and 2 can reconstruct the undirected skeleton of Bayesian networks of varying variable numbers and densities, with a much greater fidelity than a random approach. However, they cannot compete with traditional structure learning approaches in this regard, as they do not recover edge directions or conditional probability distributions. That is not to say that these algorithms are useless in this regard. Indeed, a maximum weighted spanning tree approach followed by directing edges away from a given node (as described in Section III-A), may prove to be a useful initial condition in traditional Bayesian network structure learning methods, although further research is required to test this hypothesis.

The real strength of these algorithms, however, is in knowledge discovery and data exploration, as they provide a quick and computationally cheap method to graphically represent influence within systems of random variables, to a reasonable degree of accuracy. These explorations may prove useful in understanding the density of the underlying ontology and its structural characteristics (for example, uncovering the six branches of the CHILD Network in Figure 9).

While this research has focused entirely on recovering Bayesian network skeletons, it can, in principle, be applied to any graphical structure modelling an influence ontology. For example, consider the problem of student plagiarism, in which a class of students produce some work and some subset of these students influence each other's work (e.g. copying). Plagiarism detection software exists that can quantify the similarity between pairs of submissions, and therefore, using this pairwise similarity as our similarity measure, we can apply algorithm 1 to discover cliques of students who worked in groups. Further research is necessary to determine the degree to which these algorithms can recover these types of undirected, highly cyclic and cliqued influence structures.

Similar work can be done to cluster individual data points based on a distance metric. For example, pieces of music could be clustered using algorithm 1 with the distance between content-based audio features as the similarity function, the hope being that these clusters reveal some property inherent to the pieces of music (for example, genre or artist). Once again, further research is required to investigate this application.

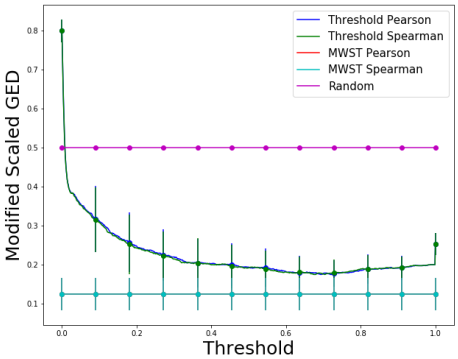
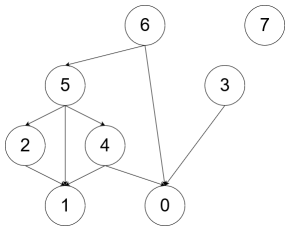
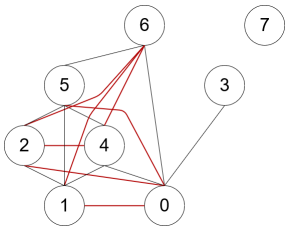


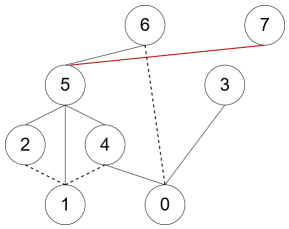
Fig. 4: Modified Scaled GED Score vs Threshold, $\rho = 0.2$, $N = 20$



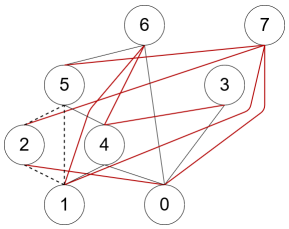
(a) *Ground truth* Bayesian network structure



(b) Threshold Reconstruction $t = 0.15$, $GED = 7$ (0.25 scaled)



(c) MWST Reconstruction, $GED = 4$ (0.1429 scaled)



(d) Random Reconstruction, $GED = 11$ (0.3929 scaled)

Fig. 5: Example reconstructions for a graph where $N = 8$ and $\rho = 0.4$. Black lines denote correctly reconstructed edges, thick red lines denote incorrectly added edges and dotted lines denote edges that should have been added but were not

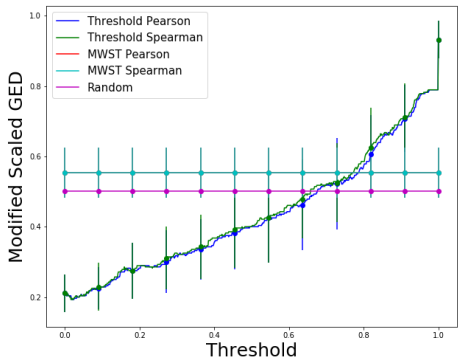


Fig. 6: Modified Scaled GED Score vs Threshold, $\rho = 0.8$, $N = 8$

References

[1] D. Justice and A. Hero, "A linear formulation of the graph edit distance for graph recognition," *IEEE Trans. on Pattern Analysis and Machine*

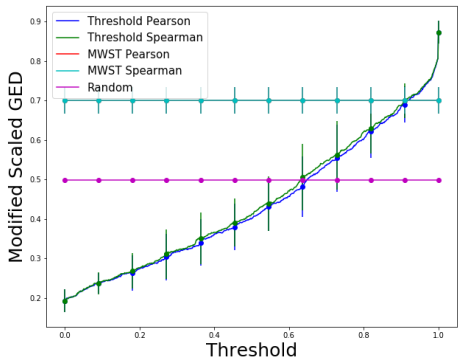


Fig. 7: Modified Scaled GED Score vs Threshold, $\rho = 0.8$, $N = 16$

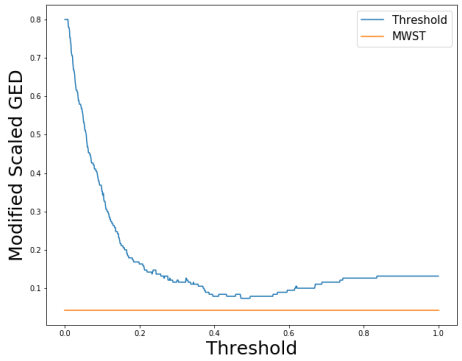
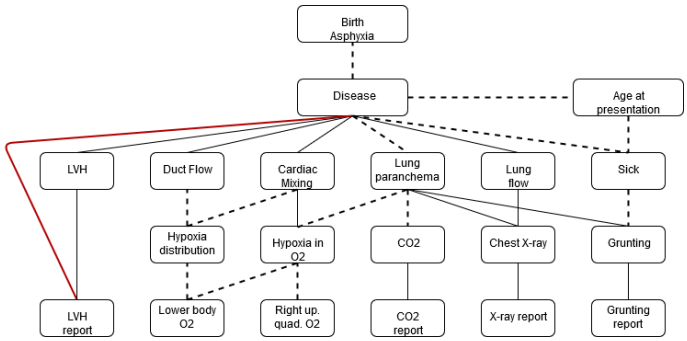
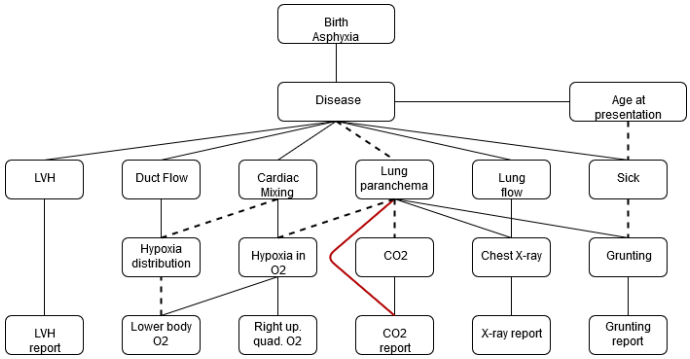


Fig. 8: GED vs. t for both algorithms with bias-corrected Cramèr's V as a similarity metric



(a) Threshold reconstruction with $t = 0.4715$, $GED = 14$ (0.07368 scaled)



(b) MWST reconstruction, $GED = 8$ (0.042105 scaled)

Fig. 9: Best reconstructions of the CHILD Network. Black lines denote correctly reconstructed edges, thick red lines denote incorrectly added edges and dotted lines denote edges that should have been added but were not

Intelligence (PAMI), 2005.

[2] D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen, R. G. Cowell *et al.*, "Bayesian analysis in expert systems," *Statistical science*, vol. 8, no. 3, pp. 219–247, 1993.

[3] T. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *International Journal Human-Computer Studies*, vol. 43, pp. 907–928, 1993.

[4] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.

[5] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," *Machine Learning*, vol. 20, no. 3, pp. 197–243, 1995.

[6] R. Ajoodha and B. Rosman, "Tracking influence between naïve Bayes models using score-based structure learning," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*. IEEE, Nov. 2017.

[7] X.-w. Chen, G. Anantha, and X. Wang, "An effective structure learning method for constructing gene networks," *Bioinformatics*, vol. 22, no. 11, pp. 1367–1374, 2006.

[8] N. Meinshausen, P. Bühlmann *et al.*, "High-dimensional graphs and variable selection with the lasso," *The annals of statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.

[9] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[10] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[11] D. Koller and N. Friedman, *Probabilistic Graphical Models - Principles and Techniques*. MIT Press, 2009.

[12] P. Norvig and S. J. Russell, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1994.

[13] T. Verma and J. Pearl, "Equivalence and synthesis of causal models," in *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, ser. UAI '90. New York, NY, USA: Elsevier Science Inc., 1991, pp. 255–270.

[14] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical society*, vol. 7, no. 1, pp. 48–50, 1956.

[15] A. Sanfeliu and K.-S. Fu, "A distance measure between attributed relational graphs for pattern recognition," *IEEE transactions on systems, man, and cybernetics*, no. 3, pp. 353–362, 1983.

[16] E. . A. Nordhaus and J. W. Gaddum, "On complementary graphs," *The American Mathematical Monthly*, vol. 63, no. 3, pp. 175–177, 1956.

[17] W. J. C. W. H. Cunningham and W. R. P. A. Schrijver, "Combinatorial optimization," 1997.

[18] T. Rastad and N. Delfan, "The relation between topological ordering and adjacency matrix in digraphs," 2012.

[19] K. Pearson, "Notes on regression and inheritance in the case of two parents," in *Proceedings of the Royal Society of London*, vol. 58, 1895, pp. 240–242.

[20] C. Spearman, "The proof and measurement of association between two things," *American journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[21] D. J. Spiegelhalter, "Learning in probabilistic expert systems," *Bayesian statistics*, vol. 4, pp. 447–465, 1992.

[22] W. Bergsma, "A bias-correction for Cramér's V and Tschuprow's T," *Journal of the Korean Statistical Society*, vol. 42, no. 3, pp. 323–328, 2013.