Fast Constraint-Based Structure Learning for Continuous-Time Bayesian Networks

Alessandro Bregoli A.BREGOLI1@CAMPUS.UNIMIB.IT
Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Milan, Italy
Marco Midali
Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Milan, Italy
Marco Scutari
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland
Fabio Stella
Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Milan, Italy
Alessio Zanga
Department of Informatics, Systems and Communication (DISCo), University of Milano-Bicocca, Milan, Italy
Whysen S.r.l., Bellinzago Lombardo, Italy

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Continuous-time Bayesian networks (CTBNs) are a class of probabilistic graphical models representing complex processes evolving over time. Applying Bayesian networks' state-space factorisation to the time-space mathematical structure of continuous-time Markov processes (CTMPs) significantly reduces the number of parameters required to model them. However, learning the structure of a CTBN that correctly factorises the corresponding CTMP is difficult, as the set of candidate graphs grows exponentially with the size of the CTMP random vector. In this work, we present HITON-CTPC, a novel constraint-based algorithm to navigate the search space of CTBNs more efficiently. Specifically, we incorporate the key steps of the HITON-PC BN structure-learning algorithm into the CTPC CTBN algorithm, thereby enabling us to add heuristics for determining the order of conditional independence tests. We assessed the performance of the proposed solution through an extensive set of synthetic experiments spanning a wide range of state-space and time-space configurations.We observed that HITON-CTPC achieves competitive structural accuracy while significantly reducing execution time. Furthermore, we evaluated the new HITON-CTPC algorithm using the Beijing Multi-Site Air Quality data set (Chen, 2017) to assess its ability to reconstruct significant relationships in a real-world scenario.

Keywords: Continuous-time Bayesian networks, Structure learning, Causal discovery

© 2026 A. Bregoli, M. Midali, M. Scutari, F. Stella & A. Zanga.

1. Introduction

Continuous-time Bayesian networks (CTBNs) are models that learn from sparse, possibly incomplete, event-based data, providing valuable insights into the behaviour of a dynamic system. At its core, a CTBN represents a set of probability distributions that encode information about when and where the system will change its state. Recent applications can be found in multiple fields, such as medicine (Schupbach et al., 2023; Hernández Guillamet et al., 2025; Zanga et al., 2026), healthcare (Faruqui et al., 2021) and engineering (Guo et al., 2022; Shunqi et al., 2023; Liu et al., 2023), where explainability and transparency are non-negotiable requirements for high-stakes decision-making.

The first algorithm for learning a CTBN from data was proposed in Nodelman et al. (2003). We refer to this algorithm as the Continuous-Time Hill-Climbing (CTHC) algorithm. Recent work modified CTHC by exploiting the behaviour of the score function under different conditions (Acerbi et al., 2014; Linzner and Koeppl, 2018; Cąkała et al., 2024).

A parallel stream of algorithms was developed based on conditional independence testing, as for the Continuous-Time Peter-Clark (CTPC) algorithm (Bregoli et al., 2021) and its extensions (Villa-Blanco et al., 2021; Villa-Blanco et al., 2023; Zanga et al., 2026).

Most of these solutions were inspired by related work on BNs that was subsequently adapted to CTBNs. In this context, looking for approximations and shortcuts to speed up structure learning is a well-explored research area, given that the number of candidate structures grows exponentially with the number of variables.

Historically, heuristics for BNs have been developed by either approximating the score function (Scanagatta et al., 2015; de Campos et al., 2018; Guo and Constantinou, 2020) or constraining the search space (Aliferis et al., 2003). While score approximations often rely on properties specific to the probability distributions within the BNs, constraint-based heuristics are more flexible and are similar to existing CTBN learning algorithms. Yet, to the best of our knowledge, no heuristic has been proposed for CTBNs so far.

Our main contribution is a novel heuristic-based structure learning algorithm for CTBNs, in which the order of the conditional independence tests is chosen to minimise the total number of tests performed. Empirical evidence from extensive experiments on both synthetic and real-world data shows that our approach effectively detects the most likely independence statements earlier, thereby enabling faster discovery of the CTBN's structure.

2. Preliminaries

2.1. Continuous-Time Markov Processes

Definition 1 (Trajectory)
Let $\mathbf{X}$ *be a random vector. A trajectory* $\sigma : \mathbb{R}^+ \to \mathcal{D}(\mathbf{X})$ *is a right-continuous, piece-wise constant function that assigns a value to* $\mathbf{X}^t$ *for each time* $t \in \mathbb{R}^+$.

Definition 2 (Continuous-Time Markov Process)
A Continuous-Time Markov Process (CTMP; Norris, 1997) is a stochastic process over $\mathbf{X}$ *defined by a tuple* $\langle P^0_{\mathbf{X}}, \mathbf{Q}_{\mathbf{X}} \rangle$*, where:*

- $P^0_{\mathbf{X}}$ *is an* initial distribution *of the state of the process at time* $t = 0$*;*
- $\mathbf{Q}_{\mathbf{X}}$ *is an* intensity matrix *(IM) that describes the process at time* $t > 0$*.*

The IM $\mathbf{Q}_{\mathbf{X}}$ is a square matrix where:

- the *sojourn time* in state i follows an exponential distribution $\text{Exp}(q_i)$ with parameter $q_i = -q_{ii}$ and an *expected transition time* of $1/q_i$;
- the *probability of transitioning* from state i to state j follows a categorical distribution $\text{Cat}(\boldsymbol{\theta}_i)$ with parameters $\theta_{ij} = q_{ij}/q_i$ for $j \neq i$, zero otherwise.

Parameter Learning. Given a complete trajectory σ, we can estimate $\mathbf{Q}_{\mathbf{X}}$ from its sufficient statistics $\mathbf{N}_{ij}$, the number of transitions from state i to state j, and $\mathbf{T}_i$ the time spent in state i. Then, we can specify a non-informative gamma prior Γ^0 with parameters α_{ij} and τ_{ij} to derive the posterior parameters:

$$\hat{q}_{ij} = \begin{cases} (\mathbf{N}_{ij} + \alpha_{ij}) \,/\, (\mathbf{T}_i + \tau_{ij}) & i \neq j \\ -\sum_{i \neq j} (\mathbf{N}_{ij} + \alpha_{ij}) \,/\, (\mathbf{T}_i + \tau_{ij}) & i = j \end{cases}. \tag{1}$$

2.2. Continuous-Time Bayesian Networks

The size of $\mathbf{Q}_{\mathbf{X}}$ grows exponentially with the number of random variables in $\mathbf{X}$. Continuous-time Bayesian networks have been introduced to mitigate this issue.

Definition 3 (Continuous-Time Bayesian Network)
Let $\mathbf{X}$ *be a random vector, a Continuous-Time Bayesian Network (CTBN; Nodelman et al., 2002; Shelton and Ciardo, 2014) is a stochastic process over* $\mathbf{X}$ *defined by a tuple* $\langle P^0_{\mathbf{X}}, \mathcal{G}, \mathbf{Q}_{\mathbf{X}} \rangle$*:*

- $P^0_{\mathbf{X}}$ *is an initial distribution;*
- $\mathcal{G}$ *is a directed graph* $(\mathbf{V}, \mathbf{E})$*, where each* $V_i \in \mathbf{V}$ *is mapped to a* $X_i \in \mathbf{X}$*;*
- $\mathbf{Q}_{\mathbf{X}}$ *is an IM factorized into multiple* conditional *intensity matrices (CIMs)* $\mathbf{Q}_{X_i|\mathbf{\Pi}_i}$ *according to the graph* $\mathcal{G}$ *(Nodelman et al., 2002), where* $\mathbf{\Pi}_i$ *is the parent set of node* i*.*

Since $\mathcal{G}$ factorises the state space, $\mathcal{G}$ is said to be a *structure* over $\mathbf{X}$. For this reason, CTBNs are also referred to in the literature as *structured* (or *factorised*) CTMPs, which means that instead of modelling $\mathbf{Q}_{\mathbf{X}}$ we can use a collection of CIMs as $\left\{\mathbf{Q}_{X_0|\mathbf{\Pi}_0}, \ldots, \mathbf{Q}_{X_{n-1}|\mathbf{\Pi}_{n-1}}\right\}$, reducing the asymptotic space complexity.

Parameter Learning. Estimating a CIM $\mathbf{Q}_{X_i|\mathbf{\Pi}_i}$ of a CTBN is essentially equivalent to estimating the IM $\mathbf{Q}$ of a CTMP: we can apply Equation (1) to the *conditional* counts $\mathbf{N}_{ijk}$ and times $\mathbf{T}_{ij}$ along the conditioning axis.

2.3. Learning the Structure of a CTBN

While learning the parameters $\mathbf{Q}_{X_i|\mathbf{\Pi}_i}$ of a CTBN is relatively simple, learning its structure $\mathcal{G}$ is more complicated. In general, to estimate the graph $\mathcal{G}$ of a CTBN from a trajectory σ, there are two approaches: *score-based* and *constraint-based*.

The first proposed score-based algorithm is Continuous-Time Hill-Climbing (CTHC) (Nodelman et al., 2003), which learns $\mathcal{G}$ by finding the graph that maximises the log-likelihood:

$$\mathcal{G}^* = \arg\max_{\mathcal{G}} \sum_{X_k \in \mathbf{X}} \log \mathcal{L}\left(\mathbf{Q}_{X_k|\mathbf{\Pi}_k}\right) = \arg\max_{\mathcal{G}} \sum_{X_k \in \mathbf{X}} \log \mathcal{L}\left(\mathbf{q}_{X_k|\mathbf{\Pi}_k}\right) + \log \mathcal{L}\left(\boldsymbol{\theta}_{X_k|\mathbf{\Pi}_k}\right) \tag{2}$$

given the marginal likelihoods of $\mathbf{q}$ and $\boldsymbol{\theta}$:

$$\mathcal{L}\left(\mathbf{q}_{X_k|\mathbf{\Pi}_k}\right) = \prod_{\boldsymbol{\pi}_l \in \mathbf{\Pi}_k} \prod_{x_i \in X_k} \frac{\Gamma\left(\alpha_{x_i|\boldsymbol{\pi}_l} + \mathbf{N}_{x_i|\boldsymbol{\pi}_l} + 1\right) \left(\tau_{x_i|\boldsymbol{\pi}_l}\right)^{\alpha_{x_i|\boldsymbol{\pi}_l}+1}}{\Gamma\left(\alpha_{x_i|\boldsymbol{\pi}_l} + 1\right) \left(\tau_{x_i|\boldsymbol{\pi}_l} + \mathbf{T}_{x_i|\boldsymbol{\pi}_l}\right)^{\alpha_{x_i|\boldsymbol{\pi}_l}+\mathbf{N}_{x_i|\boldsymbol{\pi}_l}+1}},$$

$$\mathcal{L}\left(\boldsymbol{\theta}_{X_k|\mathbf{\Pi}_k}\right) = \prod_{\boldsymbol{\pi}_l \in \mathbf{\Pi}_k} \prod_{x_i \in X_k} \frac{\Gamma\left(\alpha_{x_i|\boldsymbol{\pi}_l}\right)}{\Gamma\left(\alpha_{x_i|\boldsymbol{\pi}_l} + \mathbf{N}_{x_i|\boldsymbol{\pi}_l}\right)} \prod_{x_j \in X_k} \frac{\Gamma\left(\alpha_{x_i x_j|\boldsymbol{\pi}_l} + \mathbf{N}_{x_i x_j|\boldsymbol{\pi}_l}\right)}{\Gamma\left(\alpha_{x_i x_j|\boldsymbol{\pi}_l}\right)}.$$

Maximising the log-likelihood directly may lead to overfitting. Therefore, it is common practice to add a penalty proportional to the number of parameters $|\mathbf{Q}_{X_k|\mathbf{\Pi}_k}|$ or the number of samples $|\mathcal{D}|$, as in the Bayesian Information Criterion (BIC; Schwarz, 1978):

$$BIC\left(\mathbf{Q}_{X_k|\mathbf{\Pi}_k}\right) = \log \mathcal{L}\left(\mathbf{Q}_{X_k|\mathbf{\Pi}_k}\right) - \frac{1}{2}|\mathbf{Q}_{X_k|\mathbf{\Pi}_k}| \log |\mathcal{D}|\,.$$

The first constraint-based algorithm was introduced by Bregoli et al. (2021) as the Continuous-Time Peter-Clark (CTPC) algorithm, an adaptation of the original PC algorithm for BNs. It starts from a complete graph and performs conditional independence testing to iteratively remove edges between independent variables until no more edges can be removed. To do so, we need to test independence with respect to both states and times, namely, the state-to-state transition and the time-to-transition hypotheses.

Definition 4 (Null State-To-State-Transition Hypothesis) *Given X_i, X_j and the set $\mathbf{S} \subseteq \mathbf{X} \setminus \{X_i, X_j\}$, the null state-to-state-transition hypothesis of X_j over X_i is:*

$$H_0^S : \theta_{x_i|x_j,\mathbf{s}} = \theta_{x_i|\mathbf{s}} \qquad \forall x_i \in X_i, x_j \in X_j, \mathbf{s} \in \mathbf{S}\,;$$

Definition 5 (Null Time-To-Transition Hypothesis) *Given X_i, X_j and the set $\mathbf{S} \subseteq \mathbf{X} \setminus \{X_i, X_j\}$, the null time-to-transition hypothesis of X_j over X_i is:*

$$H_0^T : q_{x_i|x_j,\mathbf{s}} = q_{x_i|\mathbf{s}} \qquad \forall x_i \in X_i, x_j \in X_j, \mathbf{s} \in \mathbf{S}\,.$$

The null state-to-state-transition hypothesis can be tested using the χ^2 test as:

$$\chi^2_{|X|-1} = \sum_{x'_i \in X_i} \frac{(K \cdot \mathbf{N}_{x_i x'_i|x_j,\mathbf{s}} - L \cdot \mathbf{N}_{x_i x'_i|\mathbf{s}})^2}{\mathbf{N}_{x_i x'_i|x_j,\mathbf{s}} + \mathbf{N}_{x_i x'_i|\mathbf{s}}}, \quad K = \sqrt{\frac{\sum_{x'_i \in X_i} \mathbf{N}_{x_i x'_i|\mathbf{s}}}{\sum_{x'_i \in X_i} \mathbf{N}_{x_i x'_i|x_j,\mathbf{s}}}}, \quad L = \frac{1}{K},$$

while the null time-to-transition hypothesis can be tested via an F-test as:

$$F_{r_1,r_2} = \frac{q_{x_i|\mathbf{s}}}{q_{x_i|x_j,\mathbf{s}}}, \quad r_1 = \sum_{x'_i \in X_i} \mathbf{N}_{x_i x'_i|x_j,\mathbf{s}}, \quad r_2 = \sum_{x'_i \in X_i} \mathbf{N}_{x_i x'_i|\mathbf{s}}.$$

Note that, under complete observability, there exists one and only one structure that represents the process factorised by the CTBN, unlike in the BNs, where it may be the case that multiple structures could represent the same probability distribution.

2.4. Heuristics for Constraint-based Approaches

Compared to score-based approaches, heuristics for constraint-based algorithms are generally less studied. Spirtes et al. (2003) report the three-step heuristic for choosing the order of the conditional independence tests in the PC algorithm:

1. Test $X_i \perp\!\!\!\perp X_j \mid \mathbf{S}$ in lexicographical order.
2. For each $X_i \perp\!\!\!\perp X_j \mid \mathbf{S}$, first test the (X_i, X_j) pairs that are least dependent in probability, then test the conditioning subsets $\mathbf{S}$ in lexicographical order.
3. For each variable X_i, first test those variables X_j that are least dependent in probability on X_i, conditional on the $\mathbf{S}$ that are most dependent in probability on X_j.

The idea underlying these three heuristics is that the faster the edges are eliminated, the smaller the search space becomes at later stages of the algorithm, and the faster the algorithm runs. Hence, it is best to select first for testing those variable pairs X_i and X_j and subsets $\mathbf{S}$ for which X_i and X_j are most likely to be conditionally independent given $\mathbf{S}$.

The HITON-PC algorithm (Aliferis et al., 2003) builds on the same intuition encoded in this heuristic. It starts by iterating over all the X_i, estimating a measure of (unconditional) association ϕ between X_i and the candidates' neighbours $\mathbf{X} \setminus \{X_i\}$. It then selects the X_j most associated with X_i, which matches the second part of step 3, where we want to test for the least dependent pair (X_i, X_j) conditionally on $\mathbf{S}$. The pseudocode is reported in Algorithm 1.

3. A Heuristic-Based Structure Learning Algorithm for CTBNs

In this work, we start from the idea proposed in Aliferis et al. (2003) to develop a novel constraint-based algorithm that uses a heuristic to significantly reduce the execution time of structure learning while achieving performance comparable to the original CTPC.

Algorithm 1: HITON-PC

Input: Data set $\mathcal{D}$, target variable X_i, a measure of association $\phi(\cdot, X_i)$.
Output: Candidate parents and children of target variable X_i.

$\mathbf{C} \leftarrow \varnothing$; // Initialize the resulting set.
$\mathbf{Q} \leftarrow \mathbf{X} \setminus \{X_i\}$; // Initialize the to-be-visited set.
repeat
  $X_j \leftarrow \arg\max_{X_k \in \mathbf{Q} \setminus \mathbf{C}} \phi(X_k, X_i)$; // Find $X_j \notin \mathbf{C}$ that maximizes ϕ.
  $\mathbf{C} \leftarrow \mathbf{C} \cup \{X_j\}$; // Admit X_j into $\mathbf{C}$.
  foreach $X_j \in \mathbf{C}$ **do** // If there is a variable X_j ...
    foreach $\mathbf{S} \subseteq \mathbf{C} \setminus \{X_j\}$ **do** // ... and a subset $\mathbf{S} \subseteq \mathbf{C}$ ...
      if $X_j \perp\!\!\!\perp X_i \mid \mathbf{S}$ **then** // ... s.t. $X_j \perp\!\!\!\perp X_i \mid \mathbf{S}$.
        $\mathbf{C} \leftarrow \mathbf{C} \setminus \{X_j\}$; // Remove X_j from $\mathbf{C}$.
        $\mathbf{Q} \leftarrow \mathbf{Q} \setminus \{X_j\}$; // Do not consider X_j again.
        break;
until $\mathbf{Q} \setminus \mathbf{C} = \varnothing$; // No more variables are left.
return $\mathbf{C}$

We modified the HITON algorithm to work with the CTBN framework. Specifically, we defined a heuristic for ordering candidate parents and then incorporated it into HITON-PC to learn the network structure, exploiting the peculiarities of the CTBN framework. Similar to the adaptation of the PC algorithm for the CTBN, some procedures of the original HITON algorithm, such as HITON-MB, can be removed because conditional independencies are not symmetric under complete observability; therefore, searching the Markov blanket is not necessary to identify the network structure.

3.1. Heuristic Quantification of Marginal Dependence

The first step in implementing a version of the HITON algorithm for CTBNs is to identify an effective heuristic to quantify the strength of the marginal dependence between two nodes. We decided to use the p-values from the hypothesis tests assessing independence among variables. Indeed, different tests have a different interpretation for the p-value, thus we used the following transformation to interpret each p-value ρ as if it came from a right-sided test:

$$\rho^* = \begin{cases} \rho & \text{If } \rho \text{ comes from a right-tailed test,} \\ 1 - \rho & \text{If } \rho \text{ comes from a left-tailed test,} \\ 2 \cdot |\rho - 0.5| & \text{If } \rho \text{ comes from a two-tailed test.} \end{cases} \tag{3}$$

Furthermore, identifying marginal independence between two nodes within the CTBN framework may require multiple hypothesis tests, some of which are applied repeatedly. To aggregate the resulting p-values into a single number, we compute the mean $\overline{\rho^*}$ of all p-values ρ^* generated during the process.

3.2. The HITON-CTPC Algorithm

The main idea behind the original HITON-PC (Algorithm 1) is to optimise the sequence of hypothesis tests to perform as few of them as possible. With our adaptation, we make a step further; specifically, we trade lower reconstruction accuracy for shorter execution time. To achieve this goal, we evaluate only the subset of conditional independence $X_i \perp\!\!\!\perp X_j | \mathbf{S}_{ij}$ where the separation set $\mathbf{S}_{ij}$ is defined as follows:

$$\mathbf{S}_{ij} = \left\{ X_k \in \mathbf{X} : \begin{array}{c} \overline{\rho^*}_{ij} < \overline{\rho^*}_{ik} \\ X_i \not\perp\!\!\!\perp X_k | \mathbf{S}_{ik} \end{array} \right\}, \tag{4}$$

where $\overline{\rho^*}_{ij}$ is the transformed mean p-value computed by the hypothesis test used to determine the marginal independence between the child variable X_i and the parent variable X_j. Informally, we employ a heuristic where X_j is tested for independence from node X_i by conditioning only on nodes with a higher $\overline{\rho^*}$ than X_j. Algorithm 2 formalises the proposed adaptation of the HITON algorithm to CTBNs. Line 1 initialises the candidate parent set with all the variables in $\mathbf{X}$ excluding the target (child) variable X_i.

Algorithm 2: HITON-CTPC

Input: Data set $\mathcal{D}$, target variable X_i.
Output: Parent set for target variable X_i.

```
1  C ← X \ {X_i} ;                          // Initialize the candidate parents set.
2  foreach X_j ∈ C do
3  |  if X_i ⫫ X_j then                                  // Test for X_i ⫫ X_j | ∅.
4  |  |  C ← C \ {X_j} ;                          // Do not reject null hypothesis.
5  |  else
6  |  |  ρ[X_j] ← ρ̄*_ij ;                          // Store the mean of ρ* for X_j.
7  S ← ∅ ;                                  // Initialize the resulting parents set.
8  C ← arg sort_{X_j ∈ C} ρ[X_j] ;              // Sort C by ρ in descending order.
9  foreach X_j ∈ C do
10 |  if X_i ⫫̸ X_j | S then
11 |  |  S ← S ∪ {X_j};
12 return S
```

Lines 2 through 6 evaluate the independence of X_i from each variable X_j in the candidate parent set. If $X_i \perp\!\!\!\perp X_j$ then X_j is removed from the parent set, otherwise, the $\overline{\rho^*}_{ij}$ is stored. Next, lines from 9 to 11 iterate over each remaining node X_j in descending order of their $\overline{\rho^*}$. For each variable X_j, if the conditional dependence $X_i \not\perp\!\!\!\perp X_j \mid \mathbf{S}$ holds, X_j is added to the parent set. Finally, line 12 returns the finalised parent set $\mathbf{S}$ for node X_i. The main difference between the algorithm we propose and the original algorithm is that HITON-PC defines a new order for the hypothesis tests, whereas our algorithm also limits the space of possible tests. Furthermore, our algorithm tests the conditional independence using all the nodes already inserted into the parent set (Algorithm 2 as a separating set:11). In contrast, the original HITON-PC tests the conditional independence for each possible subset of the candidate parent set (Algorithm 1:7).

4. Experiments

This section provides an empirical validation of the HITON-CTPC algorithm. Because the proposed method relies on a heuristic, it is essential to verify that its ability to identify the correct structure remains intact. Furthermore, we need to evaluate the improvement in execution time against CTPC and CTHC. We decided to implement all the structure learning and parameter learning algorithms used in the experiment from scratch to fairly compare their performance, especially in terms of execution time. The implementation is written using a combination of Rust and Python. The rest of this section is organised as follows: Section 4.1 describes a series of simulation experiments to benchmark performance in a controlled environment with known ground truth; Section 4.2 describes the application of HITON-CTPC to a real-world data set, comparing the relationships identified by our algorithm with knowledge elicited from the literature.

4.1. Synthetic Experiments

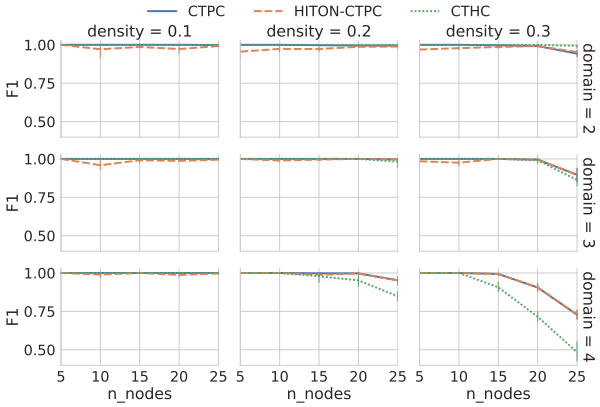
The synthetic experiments aim to assess HITON-CTPC's performance in terms of both structure reconstruction and execution time. To this end, we randomly generated the structures and parameters for a variety of CTBNs and sampled a set of trajectories for each. This generation procedure was repeated 5 times for each possible combination of parameters listed in Table 1, following these steps: *(i)* create an empty network, *(ii)* randomly add edges up to the selected network density, *(iii)* randomly generate the parameters according to the structure of the network where: $\mathbf{q}_i$ is sampled from a uniform distribution between *Min Intensity* and *Max Intensity*, θ_i is generated from a uniform distribution subject to the constraints for the rows of a IM (Equation 1). Then, for each network, we generated 100 trajectories using the forward sampling described in Nodelman (2007). For each trajectory, we generate a number of events proportional to the number of nodes by using the following formula: Number of events = 100000 × Nodes, which allows experiments to have a similar number of transitions for each node.

The structure learning task can be viewed as a binary classification problem in which the goal is to predict the presence/absence of each possible edge. Thus, as in Bregoli et al. (2021), we use the F1 score to assess the reconstruction accuracy of the three algorithms. Furthermore, we employ the Relative Structural Hamming Distance (RSHD), defined as the standard SHD normalized by the total number of possible edges.

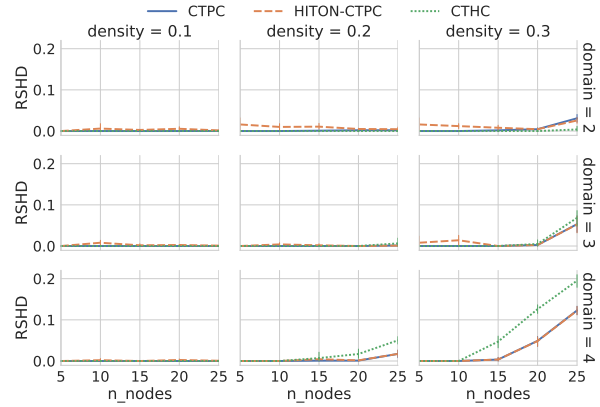
Parameter	Values
Number of Nodes	5, 10, 15, 20, 25
Network Density	0.1, 0.2, 0.3
Node Cardinality	2, 3, 4

Parameter	Values
Min Intensity	2.0
Max Intensity	5.0

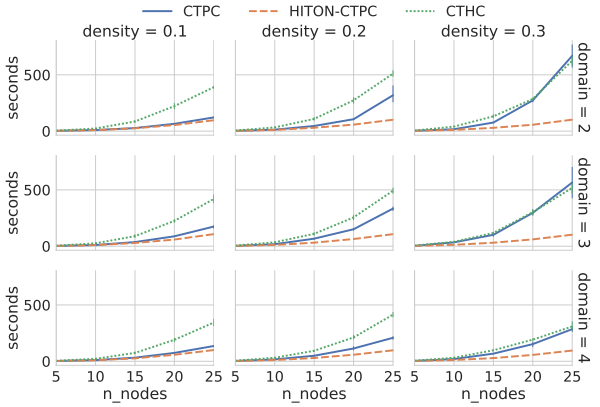
Table 1: Parameters of the simulation experiments. Network Density: 0.0 means no edges, whereas 1.0 means all the possible edges. Node Cardinality: the number of values that all nodes in a CTBN can assume. Min Intensity: minimum value generated for a CIM's main diagonal. Max Intensity: maximum value generated for a CIM's main diagonal.



(*a*) F1



(*b*) RSHD



(*c*) Execution Time

Figure 1: Performance comparison of CTPC, HITON-CTPC and CTHC on the synthetic data set. (a) Reconstruction accuracy measured by F1 score. (b) Reconstruction accuracy measured by RSHD (c) Execution time in seconds.

Figures 1(a), 1(b) and 1(c) show that HITON-CTPC achieves comparable performance to CTPC while requiring less time to identify the structure. To statistically validate this runtime advantage, we conducted a one-sided Wilcoxon signed-rank test. The results demonstrate that HITON-CTPC achieves a statistically significant reduction in execution time compared to CTPC and CTHC ($\alpha = 0.01$). We applied the same statistical framework to evaluate the algorithms' accuracy metrics, specifically comparing their F1-scores, false positives (FP), and false negatives (FN). A one-sided Wilcoxon signed-rank test indicates that CTPC attains a significantly higher F1-score ($\alpha = 0.01$) than both CTHC and HITON-CTPC, while the difference between CTHC and HITON-CTPC is not statistically significant ($\alpha = 0.01$). However, a deeper analysis of the error types reveals distinct trade-offs: HITON-CTPC exhibits a statistically higher FP rate, while CTHC and CTPC display statistically higher FN rates. This behaviour is explained by the algorithmic design of HITON-CTPC, which uses a heuristic to limit the number of evaluated conditional independence tests. Consequently, it is more likely to omit the specific tests required to detect conditional independence and remove redundant edges. For the same reason, since HITON-CTPC evaluates fewer tests, the overall type-II error (FN) is lower.

Finally, in the synthetic experiments with the highest density (0.3) and the largest node cardinality (4), all models exhibit a marked performance deterioration. For these networks, the generated data set is too small to retrieve all the edges of the real structure effectively. Hence, the execution time is lower than in the same experiments conducted at lower density or cardinality because the models incorrectly identify conditional independence among variables, thereby introducing fewer edges and resulting in lower execution time.

4.2. Pollution Data Set

This real-world experiment aims to validate the HITON-CTPC algorithm on a well-known data set and compare the identified relationships with the specialised literature. For this reason, we decided to use the Beijing Multi-Site Air Quality data set (Chen, 2017); a high-resolution collection of hourly air quality and meteorological data recorded between March 2013 and February 2017 across 12 monitoring sites in Beijing, China. For the analysis, we focused on the data collected at the Gucheng station.

To address the presence of missing values and ensure a robust time series, the data set was temporally aggregated from hourly to daily intervals, computing the mean for continuous variables and the mode for discrete variables. Then, since CTBNs natively handles only discrete data, we discretized the continuous quantities with the method described in Wu et al. (2024), refining the expert-informed thresholds to eliminate sparse tail counts and ensure statistical stability during continuous-time structure learning. The only exception was made for the rain variables, for wich we followed the standard classification of the China Meteorological Administration (CMA).

Table 2 summarises the variables available in the data set with the cardinality identified by the discretisation procedure.

Class	Variable	$\lvert X_i \rvert$
Weather	Pressure (PRES)	2
	Temperature (TEMP)	7
	Dew point (DEWP)	10
	Rain (RAIN)	4
	Wind speed (WSPM)	2
	Wind direction (wd)	8

Class	Variable	$\lvert X_i \rvert$
Pollution	Particulate Matter (PM2.5)	6
	Particulate Matter (PM10)	4
	Sulfur dioxide (SO2)	2
	Nitrogen dioxide (NO2)	3
	Carbon monoxide (CO)	3
	Ozone (O3)	2

Table 2: Variables in the Beijing Multi-Site Air Quality data set and their cardinality $|X_i|$.

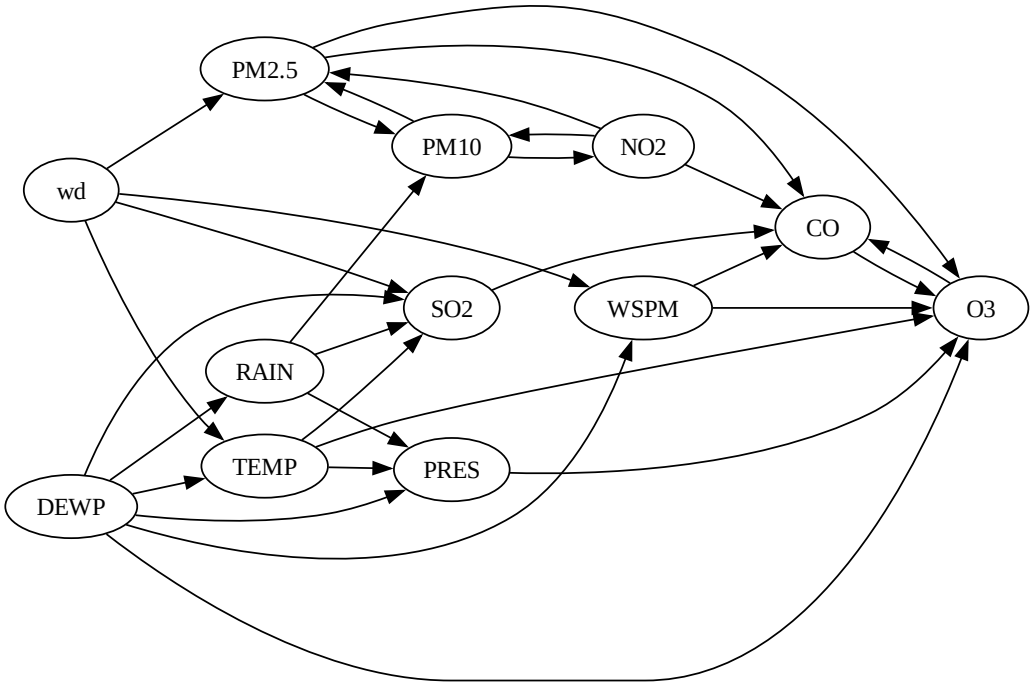


Figure 2: Structure of the CTBN learned by HITON-CTPC from the Beijing Multi-Site Air Quality data set.

We learned the network structure using the HITON-CTPC algorithm on the discretised data set. To maintain physical consistency as suggested by Zhang et al. (2017), a blacklist was enforced to prevent pollutants from influencing meteorological nodes. To ensure robust edge selection for both the F-test and the χ^2-test, we used an alpha equal to 7×10^{-7}.

The Beijing Multi-Site Air Quality data set does not have a single ground truth for the relationships among the different variables. Nevertheless, the specialised literature widely explored the interaction between pollutants and weather. Thus, the validation of the structure identified by HITON-CTPC (Figure 2) is based on the specialised literature.

Specifically, the relations, RAIN → PM10 and RAIN → SO2 represent the "washout" effect described by Zhao et al. (2016). Furthermore, the relationship DEWP → RAIN effectively captures the physical phenomena where high moisture levels precede precipitation. Finally, Lawrence (2005) establishes a robust physical relationship between DEWP and TEMP correctly identified by HITON-CTPC.

Regarding meteorological drivers, Zhang et al. (2017) observe that northerly winds typically transport colder air masses into the region, justifying the learned wd → TEMP connection. Finally, the link TEMP → SO2 reflects seasonal patterns; as noted by He et al. (2001), lower temperatures during winter lead to increased coal combustion for heating, which directly results in higher atmospheric concentrations of SO_2.

It is important to acknowledge that certain expected connections are either missing or misidentified in the learned structure. For example, the relation between PM2.5 and SO2 is completely absent, even if SO_2 is a precursor of $PM_{2.5}$ (Guo et al., 2014). Furthermore, the whashout effect should also affect PM2.5 and NO2.

Finally, note that the discretisation procedure is critical and can affect the relationships identified by the structure learning algorithms. Indeed, a poor discretisation can make a relation disappear or be reversed. It follows that some of the errors highlighted during the evaluation may depend on the discretisation rather than on the algorithm.

5. Conclusion

In this work, we introduced HITON-CTPC, a novel constraint-based structure learning algorithm for Continuous-Time Bayesian Networks (CTBNs). Inspired by the HITON framework, our approach uses a heuristic to rank the strength of relationships among variables, leveraging this ordering to significantly limit the number of hypothesis tests required during learning. Through extensive synthetic experiments, we demonstrated that HITON-CTPC maintains structural reconstruction accuracy comparable to standard CTPC while significantly reducing execution time. Finally, we validated the algorithm on a real-world data set (Chen, 2017), confirming that, even in highly complex settings such as weather-pollution systems, HITON-CTPC successfully identifies significant relationships.

Acknowledgments

Fabio Stella was supported by the MUR under the grant "Dipartimenti di Eccellenza 2023-2027" of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Milan, Italy, and by the National Plan for NRRP Complementary Investments (Project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (ANTHEM)).

References

Enzo Acerbi, Teresa Zelante, Vipin Narang, and Fabio Stella. Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell

differentiation. *BMC Bioinformatics*, 15(1):387, 12 2014. ISSN 1471-2105. doi: 10.1186/s12859-014-0387-x. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-014-0387-x.

Constantin F Aliferis, Ioannis Tsamardinos, and Alexander Statnikov. Hiton: a novel markov blanket algorithm for optimal variable selection. In *AMIA annual symposium proceedings*, volume 2003, page 21. American Medical Informatics Association, 2003.

Alessandro Bregoli, Marco Scutari, and Fabio Stella. A constraint-based algorithm for the structural learning of continuous-time Bayesian networks. *International Journal of Approximate Reasoning*, 138:105–122, 11 2021. ISSN 0888613X. doi: 10.1016/j.ijar.2021.08.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0888613X21001304.

Tomasz Cąkała, Błażej Miasojedow, Wojciech Rejchel, and Maryia Shpak. Structure learning for continuous time Bayesian networks via penalized likelihood. *Scandinavian Journal of Statistics*, 51(4):1707–1729, 12 2024. ISSN 14679469. doi: 10.1111/SJOS.12747.

Song Chen. Beijing Multi-Site Air Quality. UCI Machine Learning Repository, 2017. URL https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data. DOI: https://doi.org/10.24432/C5RK5G.

Cassio P. de Campos, Mauro Scanagatta, Giorgio Corani, and Marco Zaffalon. Entropy-based pruning for learning Bayesian networks using BIC. *Artificial Intelligence*, 260:42–50, 7 2018. ISSN 00043702. doi: 10.1016/j.artint.2018.04.002. URL https://linkinghub.elsevier.com/retrieve/pii/S000437021830167X.

Syed Hasib Akhter Faruqui, Adel Alaeddini, Jing Wang, Susan P. Fisher-Hoch, and Joseph B. McCormick. Dynamic Functional Continuous Time Bayesian Networks for Prediction and Monitoring of the Impact of Patients' Modifiable Lifestyle Behaviors on the Emergence of Multiple Chronic Conditions. *IEEE Access*, 9:169092–169106, 2021. ISSN 21693536. doi: 10.1109/ACCESS.2021.3136618. URL https://ieeexplore.ieee.org/abstract/document/9656141.

Song Guo, Min Hu, Misti L. Zamora, Jianfei Peng, Dongjie Shang, Jing Zheng, Zhuofei Du, Zhijun Wu, Min Shao, Limin Zeng, Mario J. Molina, and Renyi Zhang. Elucidating severe urban haze formation in china. *Proceedings of the National Academy of Sciences*, 111(49):17373–17378, November 2014. ISSN 1091-6490. doi: 10.1073/pnas.1419604111.

Yifan Guo, Junjie Zheng, Rongjun Zhang, Youbin Yang, Yifan Guo, Junjie Zheng, Rongjun Zhang, and Youbin Yang. An evidence-based risk decision support approach for metro tunnel construction. *Journal of Civil Engineering and Management*, 28(5):377–396, 5 2022. ISSN 1822-3605. doi: 10.3846/JCEM.2022.16807. URL https://journals.vilniustech.lt/index.php/JCEM/article/view/16807.

Zhigao Guo and Anthony C. Constantinou. Approximate Learning of High Dimensional Bayesian Network Structures via Pruning of Candidate Parent Sets. *Entropy 2020, Vol. 22, Page 1142*, 22(10):1142, 10 2020. ISSN 1099-4300. doi: 10.3390/E22101142. URL https://www.mdpi.com/1099-4300/22/10/1142/htmhttps://www.mdpi.com/1099-4300/22/10/1142.

Kebin He, Fumo Yang, Yongliang Ma, Qiang Zhang, Xiaohong Yao, Chak K Chan, Steven Cadle, Tai Chan, and Patricia Mulawa. The characteristics of pm2.5 in beijing, china. *Atmospheric Environment*, 35(29):4959–4970, October 2001. ISSN 1352-2310. doi: 10.1016/s1352-2310(01)00301-6.

Guillem Hernández Guillamet, Francesc Lopez Segui, Josep Vidal-Alaball, and Beatriz Lopes. CTBN-PH: A Continuous-Time Bayesian Network for Individualised Diagnostic Risk Prediction, 2025. URL https://www.ssrn.com/abstract=5264449.

Mark Lawrence. The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, 86:225–234, 02 2005. doi: 10.1175/bams-86-2-225.

Dominik Linzner and Heinz Koeppl. Cluster Variational Approximations for Structure Learning of Continuous-Time Bayesian Networks from Incomplete Data. Proceedings of the 32nd international conference on neural information processing systems, NIPS'18, 2018.

Zheng Liu, Zhenfeng He, Liang Tu, Xin Liu, Haodong Liu, and Jinlong Liang. A fatigue reliability assessment approach for wind turbine blades based on continuous time Bayesian network and FEA. *Quality and Reliability Engineering International*, 39(5):1603–1621, 7 2023. ISSN 10991638. doi: 10.1002/QRE.3262.

Uri Nodelman. *Continuous Time Bayesian Networks*. PhD thesis, Stanford University, 2007.

Uri Nodelman, Christian R Shelton, and Daphne Koller. Continuous Time Bayesian Networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002.

Uri Nodelman, Christian R. Shelton, and Daphne Koller. Learning Continuous Time Bayesian Networks. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2003.

J. R. Norris. *Markov Chains*. Cambridge University Press, 2 1997. ISBN 9780521481816. doi: 10.1017/CBO9780511810633. URL https://www.cambridge.org/core/product/identifier/9780511810633/type/book.

Mauro Scanagatta, Cassio P. De Campos, Giorgio Corani, and Marco Zaffalon. Learning Bayesian networks with thousands of variables. In *Advances in Neural Information Processing Systems*, volume 2015-Janua, 2015.

Jordan Schupbach, Elliott Pryor, Kyle Webster, and John Sheppard. A Risk-Based Approach to Prognostics and Health Management Combining Bayesian Networks and Continuous-Time Bayesian Networks. *IEEE Instrumentation and Measurement Magazine*, 26(5):3–11, 8 2023. ISSN 19410123. doi: 10.1109/MIM.2023.10208251. URL https://ieeexplore.ieee.org/abstract/document/10208251.

Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 3 1978. ISSN 0090-5364. doi: 10.1214/aos/1176344136. URL

https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full.

Christian R. Shelton and Gianfranco Ciardo. Tutorial on structured continuous-time markov processes. *Journal of Artificial Intelligence Research*, 51:725–778, 2014. ISSN 10769757. doi: 10.1613/jair.4415.

Yang Shunqi, Zeng Ying, Li Xiang, Li Yanfeng, and Huang Hongzhong. Reliability analysis for wireless communication networks via dynamic Bayesian network. *Journal of Systems Engineering and Electronics*, 34(5):1368–1374, 10 2023. ISSN 16711793. doi: 10.23919/JSEE.2023.000130. URL https://ieeexplore.ieee.org/abstract/document/10308768.

Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, Prediction, and Search. 2nd edn. *Statistics in Medicine*, 22(13):2236–2237, 7 2003. ISSN 0277-6715. doi: 10.1002/sim.1415.

Carlos Villa-Blanco, Alessandro Bregoli, Concha Bielza, Pedro Larrañaga, and Fabio Stella. Constraint-based and hybrid structure learning of multidimensional continuous-time Bayesian network classifiers. *International Journal of Approximate Reasoning*, 159:108945, 8 2023. ISSN 0888613X. doi: 10.1016/j.ijar.2023.108945. URL https://linkinghub.elsevier.com/retrieve/pii/S0888613X23000762.

Carlos Villa-Blanco, Pedro Larrañaga, and Concha Bielza. Multidimensional continuous time Bayesian network classifiers. *International Journal of Intelligent Systems*, 36(12):7839–7866, 12 2021. ISSN 0884-8173. doi: 10.1002/int.22611. URL https://onlinelibrary.wiley.com/doi/10.1002/int.22611.

Xiaoxuan Wu, Qiang Wen, and Jun Zhu. Association rule mining with a special rule coding and dynamic genetic algorithm for air quality impact factors in beijing, china. *PLOS ONE*, 19(3):e0299865, March 2024. ISSN 1932-6203. doi: 10.1371/journal.pone.0299865.

Alessio Zanga, Francesca Graziano, Giuseppe Citerio, Paola Rebora, Stefania Galiberti, Shubhayu Bhattacharyay, David K. Menon, Ewout W. Steyerberg, and Fabio Stella. Learning a Causal Model for Intracranial Pressure in Patients with Traumatic Brain Injury. pages 270–282, 2026. ISSN 1865-0937. doi: 10.1007/978-3-032-16708-8_22. URL https://link.springer.com/chapter/10.1007/978-3-032-16708-8_22.

Shuyi Zhang, Bin Guo, Anlan Dong, Jing He, Ziping Xu, and Song Xi Chen. Cautionary tales on air-quality improvement in beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205):20170457, 2017. ISSN 1471-2946. doi: 10.1098/rspa.2017.0457.

Suping Zhao, Ye Yu, Daiying Yin, Jianjun He, Na Liu, Jianjun Qu, and Jianhua Xiao. Annual and diurnal variations of gaseous and particulate pollutants in 31 provincial capital cities based on in situ air quality monitoring data from china national environmental monitoring center. *Environment International*, 86:92–106, January 2016. ISSN 0160-4120. doi: 10.1016/j.envint.2015.11.003.