Our algorithm is going to calculte a partial time graph (so a graph that only accounts causal effects into the current time step t). We need the following assumption:

- finite time causal effects
- time invariant causal effects
- temporal-causal constraint without instantaneous effects

Because we do not know the exact maximum causal lag (called p_{max}) in advance, we want to calculate this graph incrementally over the concidered time lags p, starting with a small value for p (e.g. $p = 0$). At first we introduce the following definitions and notations:

$$\begin{aligned}
m &: \text{number of dimensions of our multivariate time series} \\
X_i^t &: \text{random variable of dimension i at time t for } 1 \leq i \leq m \\
& \text{ respectively the corresponding node} \\
G = (V, E) &: \text{original full time graph that generated the data} \\
G_p = (V_p, E_p) &: \text{generated partial time graph with time lag p} \\
Par(X) &: \text{parents of node X in graph G} \\
Par_p(X) &: Par(X) \cap V_p \text{ (parents of X, contained in } G_p) \\
In_p(X) &: \text{set of nodes with arrow to X in generated graph } G_p
\end{aligned}$$

When generating the partial time graph G_{p+1} based on G_p, there are 2 main questions

- which of the new nodes $X_i^{t-(p+1)}$ has an arrow to a node in X_j^t
- which of the arrows from any $X_i^{t-p \leq k \leq t-1}$ to any node $X_{1 \leq j \leq m}^t$ is not causal and thus can be deleted

We use the following algorithm to solve this.

Algorithm $G_p \longrightarrow G_{p+1}$**:**

1. add the new nodes to the graph $V_{p+1} = V_p \cup \{X_i^{t-(p+1)} | 1 \leq i \leq m\}$
2. draw an arrow X_i^{t-p-1} to X_j^t if $X_i^{t-p-1} \not\perp\!\!\!\perp X_j^t$
   (this gives us a premature new edge set E'_{p+1} and corresponding $In'_{p+1}(X_j^t)$)
3. for each X_j^t and $X \in In'_{p+1}(X_j^t)$:
   delete the arrow $X \to X_j^t$ if $X \perp\!\!\!\perp X_j^t \mid In'_{p+1}(X_j^t) \setminus \{X\}$
   (this finally gives us E_{p+1} and the corresponding $In_{p+1}(X_j^t)$)

As soon as p is equal to the maximum causal lag p_{max}, we have a correct partial time DAG.

Proof:
I: The first thing we can show, is that for each step p and each X_j^t, we have $Par_p(X_j^t) \subseteq In_p(X_j^t)$. This can be shown by induction.

$p = 0$:
In this case our graph G_0 only consists of nodes from $V_0 = \{X_j^t | 1 \leq j \leq m\}$ and thus $Par_0(X_j^t) = V_0 \cap Par(X_j^t) = \emptyset = In_0(X_j^t)$.

$p + 1$, given it holds for p:
In step 2 of our algorithm, we add arrows $X_i^{t-(p+1)} \to X_j^t$ for new variables $X_i^{t-(p+1)}$ if they are dependent on X_j^t. This is a superset of variables with causal effect on X_j^t. In addition we already know the direction of the arrow because of the temporal-causal constraint. So given $Par_p(X_j^t) \subseteq In_p(X_j^t)$, we have $Par_{p+1}(X_j^t) \subseteq In'_{p+1}(X_j^t)$ after step 2.
In step 3 we only remove arrows $X \to X_j^t$ for $X \in In'_{p+1}(X_j^t)$ if they are conditional independent. Because conditional independence implies that there is no causal effect, we cannot remove an arrow $X \to X_j^t$ if X is a parent of X_j^t. Thus $Par_{p+1}(X_j^t) \subseteq In_{p+1}(X_j^t)$ holds after step 3.

II: The seccond thing we want to prove, is that for a variable $X \notin Par(X_j^t)$ the set $In_p(X_j^t) \setminus \{X\}$ is sufficient to block all paths from X to X_j^t in the original full time graph G, which have their last nodes in $\{X_k^\tau | t - p \leq \tau, 1 \leq k \leq m\}$.

All paths in G with last nodes in $\{X_k^\tau | t \leq \tau, 1 \leq k \leq m\}$ are already blocked, because due to causal-temporal constraint they all contain collector nodes. In addition we do not consider them while conditioning, so we also cannot unblock them.

For all paths from X to X_j^t with last nodes in $\{X_k^\tau | t-p \leq \tau < t, 1 \leq k \leq m\}$, this last node has to be a parent of X_j^t, due to temporal-causal contraint. Thus this last node is also in $Par_p(X_j^t)$. So for $X \notin Par(X_j^t)$ these paths can be blocked by conditioning over $In_p(X_j^t) \setminus \{X\}$, because $Par_p(X_j^t) \subseteq In_p(X_j^t)$ according to I.

Property II is also pretty useful, because it say that in each iteration, we keep our graph G_p as minimal as possible, given our set of informations (defined by the number of considered time steps p).

Finally we can combine both results for the step $p = p_{max}$. By definition $Par_{p_{max}}(X_j^t) = Par(X_j^t)$ and so because of I we have

$$Par(X_j^t) \subseteq In_{p_{max}}(X_j^t) \subseteq In'_{p_{max}}(X_j^t)$$

after step 2 of the algorithm. In addition II tells us that for $X \notin Par(X_j^t)$ conditioning over $In_{p_{max}}(X_j^t) \setminus \{X\}$ is now sufficient to block all paths from X to X_j^t. Thus after step 3 of the algorithm, we have

$$In_{p_{max}}(X_j^t) = In'_{p_{max}}(X_j^t) \setminus \overline{Par(X_j^t)} = Par(X_j^t)$$

So in $G_{p_{max}}$, the nodes in X_j^t for $1 \leq j \leq m$ have only incoming arrows that are also in G and our partial time graph is correct.

q.e.d.

Discussion:

- minimality of each step is a interesting property
- no fixed p_{max} anymore (only soft stopping criterion)
- polynomial runtime in each step
- need stopping criterion (e.g. information score)
- maybe has to estimate large conditional dependencies