Continuous Time Bayesian Networks with Clocks
Supplementary/Appendix

A Proofs and Derivations

In this section, we provide more detailled derivations and proofs for the claims in the paper.

A.1 Global Rates of a CTBN

A CTBN is (partially) defined via a set of conditional rates $\lambda_n : \mathcal{X}_n \times \mathcal{U}_n \to \mathbb{R}_{\geq 0}$, unique to each process, each local state and parent state. In the following, we show that the exit rate of the global process $\lambda(\boldsymbol{x})$ decomposes into a sum of these local rates.

For this, we notice, that we can express the probabilities of transitioning within small time steps in terms of rates. For the global process, the probability of changing in a small window h conditioned on all current states can be formulated as

$$
\begin{aligned}
P\left(\bigcup_n X_n(t+h) \neq x_n \;\middle|\; \bigcap_n X_n(t) = x_n\right) &= \sum_{n=1}^{N} P\left(X_n(t+h) \neq x_n \;\middle|\; \bigcap_m X_m(t) = x_m\right) && \text{(asynchronicity)}\\
&= \sum_{n=1}^{N} P\left(X_n(t+h) \neq x_n \mid X_n(t) = x_n,\, U_n(t) = \boldsymbol{u}_n\right) && \text{(conditional independence)}\\
&= \sum_{n=1}^{N} \sum_{x' \neq x_n} P\left(X_n(t+h) = x' \mid X_n(t) = x_n,\, U_n(t) = \boldsymbol{u}_n\right)\\
&= \sum_{n=1}^{N} \lambda_n(x_n,\, \boldsymbol{u}_n)\, h + o(h). && \text{(def. local exit rates)}
\end{aligned}
$$

Subsequently, in the limit $h \to 0$ we recover $\lambda(\boldsymbol{x})$:

$$
\lambda(\boldsymbol{x}) = \lim_{h \to 0} \frac{1}{h} \left(\sum_{n=1}^{N} \lambda_n(x_n,\, \boldsymbol{u}_n)\, h + o(h) \right) = \sum_{n=1}^{N} \lambda_n(x_n,\, \boldsymbol{u}_n).
$$

A.2 Transition Probabilities of a CTSMC

Similarly to CTMC's, we can also assign rates to the associated Markov process of a CTSMC. The CTSMCs transition-matrix can, as in the case of CTMCs, be expanded in orders of h in terms of instantaneous transition rates employing information of the clock

$$
\begin{aligned}
&P\left(X(t+h) = x',\, T(t+h) \in [0,\, 0+h) \mid X(t) = x,\, T(t) \in [\tau,\, \tau+h)\right)\\
&\qquad = \lambda(x,\, \tau;\, x')\, h + o(h).
\end{aligned}
$$

Similarly we can express the probability that no state change occurs

$$
\begin{aligned}
&P\left(X(t+h) = x,\, T(t+h) \in [\tau+h,\, \tau+2h) \mid X(t) = x,\, T(t) \in [\tau,\, \tau+h)\right)\\
&\qquad = 1 - \lambda(x,\, \tau)\, h + o(h).
\end{aligned}
$$

with $\lambda(x,\, \tau) = \sum_{x' \neq x} \lambda(x,\, \tau;\, x')$. Note, that the clock's exit state, does not appear on the rhs's of the equations, since the clock transitions deterministically under knowledge of the state $X(t)$, always satisfying the relation $T(t+h) - T(t) \leq h$ for any small time window h. The clock's partially deterministic dynamics can be visualized as a straight line with slope 1 over time, as long as no state change $X(t)$ occurs. In case of a state change, however, the clock resets to zero. This explains the clocks exit state in the first equation, as during the time h a clock reset must have happened. Since the clocks reachable states are bound by the time window h, the state of the clock must be found within $[0,\, 0+h)$ after h time. In the probability of no state change occurring, the clock keeps rising and is therefore found to be within $[\tau+h,\, \tau+2h)$ after h time. Any multiple state changes within h time become less probable and finally vanish, once h approaches zero, since they scale with $o(h)$.

To obtain the transition probabilities for the state alone, we need to determine which state is to be chosen, if we know, that a change has happened. Using basic rules of probability, this leads to

$$
\begin{aligned}
&P\left(X(t+h)=x' \mid X(t+h)\neq x,\, T(t+h)\in[0,\,0+h)\,,\, X(t)=x,\, T(t)\in[\tau,\,\tau+h)\right)\\
&= \frac{P\left(X(t+h)=x',\, X(t+h)\neq x,\, T(t+h)\in[0,\,0+h)\ \mid\ X(t)=x,\, T(t)\in[\tau,\,\tau+h)\right)}{P\left(X(t+h)\neq x,\, T(t+h)\in[0,\,0+h)\ \mid\ X(t)=x,\, T(t)\in[\tau,\,\tau+h)\right)} && \text{(cond. probability)}\\
&= \frac{P\left(X(t+h)=x',\, T(t+h)\in[0,\,0+h)\ \mid\ X(t)=x,\, T(t)\in[\tau,\,\tau+h)\right)}{\lambda(x,\,\tau)\,h+o(h)} && \left(\frac{\text{absorption}}{\text{def. exit rate}}\right)\\
&= \frac{\lambda(x,\,\tau;\,x')\,h+o(h)}{\lambda(x,\,\tau)\,h+o(h)} && \left(\frac{\text{def. transition rates}}{}\right)
\end{aligned}
$$

Taking the limit gives the transition probabilities from state x to x' for a clock that shows $T(t)=\tau$

$$
\lim_{h\to 0}\left(\frac{\lambda(x,\,\tau;\,x')\,h+o(h)}{\lambda(x,\,\tau)\,h+o(h)}\right)=\frac{\lambda(x,\,\tau;\,x')}{\lambda(x,\,\tau)}
$$

A.3 Survival Function and Transition Probabilities of the augmented CTBN

A.3.1 Conditional Transition Rates

Before we can proceed to work out characteristica of the global process of the augmented CTBN, we need to define the local conditional transition rates $\lambda_n(x_n,\, \tau_n,\, \boldsymbol{u}_n;\, x_n')$. These arise naturally in the expressions derived for the global process under the condition of transition asynchronicity and the restricted coupling induced by the graph $\mathcal{G}$. We define the rates $\lambda_n(x_n,\, \tau_n,\, \boldsymbol{u}_n;\, x_n')$ by

$$
\begin{aligned}
P\left(X_n(t+h)=x_n',\, T_n(t+h)\in[0,\,0+h) \mid X_n(t)=x_n,\, T_n(t)\in[\tau_n,\,\tau_n+h)\,,\, U_n(t)=\boldsymbol{u}_n\right)\\
=\lambda_n(x_n,\,\tau_n,\,\boldsymbol{u}_n;\,x_n')\,h+o(h)
\end{aligned}
$$

and the local conditional exit rate $\lambda_n(x_n,\, \tau_n,\, \boldsymbol{u}_n)$ by

$$
\begin{aligned}
&\sum_{x_n'\neq x_n} P\left(X_n(t+h)=x_n',\, T_n(t+h)\in[0,\,0+h) \mid X_n(t)=x_n,\, T_n(t)\in[\tau_n,\,\tau_n+h)\,,\, U_n(t)=\boldsymbol{u}_n\right)\\
&= P\left(X_n(t+h)\neq x_n,\, T_n(t+h)\in[0,\,0+h) \mid X_n(t)=x_n,\, T_n(t)\in[\tau_n,\,\tau_n+h)\,,\, U_n(t)=\boldsymbol{u}_n\right)\\
&= \lambda_n(x_n,\,\tau_n,\,\boldsymbol{u}_n)\,h+o(h)
\end{aligned}
$$

from which we can conclude, that $\lambda_n(x_n,\, \tau_n,\, \boldsymbol{u}_n)=\sum_{x_n'\neq x_n}\lambda_n(x_n,\, \tau_n,\, \boldsymbol{u}_n;\, x_n')$ holds in the limit $h\to 0$.

A.3.2 Global Survival Time

In order to determine the global survival function and density of the augmented CTBN, we need to find an expression of the global exit rate. Here, the procedure above from CTBN's is helpful in combination with the definition of the transition rates for CTSMC's. The probability, that a state change of the global process occurs, meaning any of the processes change and the respective clock resets, is

$$
\begin{aligned}
&P\left(\bigcup_n X_n(t+h)\neq x_n,\, T_n(t+h)\in[0,\,0+h)\ \middle|\ \bigcap_n X_n(t)=x_n,\, T_n(t)\in[\tau_n,\,\tau_n+h)\right)\\
&=\sum_{n=1}^{N} P\left(X_n(t+h)\neq x_n,\, T_n(t+h)\in[0,\,0+h)\ \middle|\ \bigcap_m X_m(t)=x_m,\, T_m(t)\in[\tau_n,\,\tau_n+h)\right) && \text{(asynchronicity)}\\
&=\sum_{n=1}^{N} P\left(X_n(t+h)\neq x_n,\, T_n(t+h)\in[0,\,0+h) \mid X_n(t)=x_n,\, T_n(t)\in[\tau_n,\,\tau_n+h)\,,\, U_n(t)=\boldsymbol{u}_n\right) && \text{(conditional independence)}\\
&=\sum_{n=1}^{N}\lambda_n(x_n,\,\tau_n,\,\boldsymbol{u}_n)\,h+o(h)\,. && \text{(def. local exit rates)}
\end{aligned}
$$

Since this covers any process in the network, we can take the limit to obtain

$$\lambda\left(\boldsymbol{x},\,\boldsymbol{\tau}\right)=\lim_{h\to0}\frac{1}{h}\left(\sum_{n=1}^{N}\lambda_{n}\left(x_{n},\,\tau_{n},\,\boldsymbol{u}_{n}\right)h+o\left(h\right)\right)=\sum_{n=1}^{N}\lambda_{n}\left(x_{n},\,\tau_{n},\,\boldsymbol{u}_{n}\right).$$

Consequently, the complementary probability that no change occurs reads

$$\begin{aligned}
&P\left(\bigcap_{n}X_{n}\left(t+h\right)=x_{n},\,T_{n}\left(t+h\right)\in\left[\tau_{n}+h,\,\tau_{n}+2h\right)\;\middle|\;\bigcap_{n}X_{n}\left(t\right)=x_{n},\,T_{n}\left(t\right)\in\left[\tau_{n},\,\tau_{n}+h\right)\right)\\
&=1-P\left(\bigcup_{n}X_{n}\left(t+h\right)\neq x_{n},\,T_{n}\left(t+h\right)\in\left[0,\,0+h\right)\;\middle|\;\bigcap_{n}X_{n}\left(t\right)=x_{n},\,T_{n}\left(t\right)\in\left[\tau_{n},\,\tau_{n}+h\right)\right)\\
&=1-\sum_{n=1}^{N}\lambda_{n}\left(x_{n},\,\tau_{n},\,\boldsymbol{u}_{n}\right)h+o\left(h\right).
\end{aligned}$$

We now formulate the path integral corresponding to the global survival function. For this, we ask for the probability of all processes keeping their state for K time windows of length h and then perform the continuous time limit. Using the shorthand for this limit ($E\equiv\{h\to0,\,K\to\infty,\,Kh=s\}$) we can write:

$$\begin{aligned}
&\Lambda\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)\\
&=\lim_{E}\prod_{k=0}^{K}P\left(\bigcap_{n}X_{n}\left(t+h\right)=x_{n},\,T_{n}\left(t+h\right)\in\left[\tau+\left(k+1\right)h,\,\tau+\left(k+2\right)h\right)\;\middle|\;\bigcap_{n}X_{n}\left(t\right)=x_{n},\,T_{n}\left(t\right)\in\left[\tau_{n}+kh,\,\tau_{n}+\left(k+1\right)h\right)\right)\\
&=\lim_{E}\prod_{k=0}^{K}\left(1-\sum_{n}\lambda_{n}\left(x_{n},\,\tau_{n}+kh,\,\boldsymbol{u}_{n}\right)h\right)+o\left(h\right)=\prod_{0}^{s}\left(1-\sum_{n}\lambda_{n}\left(x_{n},\,\tau_{n}+\sigma,\,\boldsymbol{u}_{n}\right)\mathrm{d}\sigma\right)\qquad\text{(Volterra product integral)}\\
&=\exp\left(-\int_{0}^{s}\mathrm{d}\sigma\sum_{n}\lambda_{n}\left(x_{n},\,\tau_{n}+\sigma,\,\boldsymbol{u}_{n}\right)\right)=\exp\left(-\sum_{n}\int_{\tau_{n}}^{s+\tau_{n}}\mathrm{d}\sigma\,\lambda_{n}\left(x_{n},\,\sigma,\,\boldsymbol{u}_{n}\right)\right)\\
&=\exp\left(-\sum_{n}\int_{0}^{s+\tau_{n}}\mathrm{d}\sigma\,\lambda_{n}\left(x_{n},\,\sigma,\,\boldsymbol{u}_{n}\right)\right)\exp\left(\sum_{n}\int_{0}^{\tau_{n}}\mathrm{d}\sigma\,\lambda_{n}\left(x_{n},\,\sigma,\,\boldsymbol{u}_{n}\right)\right)\\
&=\prod_{n}\frac{\Lambda_{n}\left(s+\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)}{\Lambda_{n}\left(\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)},
\end{aligned}$$

with having identified $\Lambda_{n}\left(\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)\equiv\exp\left(-\int_{0}^{\tau_{n}}\mathrm{d}\sigma\,\lambda_{n}\left(x_{n},\,\sigma,\,\boldsymbol{u}_{n}\right)\right)$ as the local survival functions. Noticing $F\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)=1-\Lambda\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)$, we can further recover the density of the global survival time from the main paper. To do so, we differentiate the global survival time c.d.f.

$$\begin{aligned}
\frac{\mathrm{d}F\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)}{\mathrm{d}s}&=-\frac{\mathrm{d}\Lambda\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)}{\mathrm{d}s}\\
&\propto\sum_{n}\lambda_{n}\left(x_{n},\,s+\tau_{n},\,\boldsymbol{u}_{n}\right)\exp\left(-\sum_{m}\int_{0}^{s+\tau_{m}}\mathrm{d}\sigma\,\lambda_{m}\left(x_{m},\,\sigma,\,\boldsymbol{u}_{m}\right)\right)\\
&=\sum_{n}\lambda_{n}\left(x_{n},\,s+\tau_{n},\,\boldsymbol{u}_{n}\right)\prod_{m}\exp\left(-\int_{0}^{s+\tau_{m}}\mathrm{d}\sigma\,\lambda_{m}\left(x_{m},\,\sigma,\,\boldsymbol{u}_{m}\right)\right)\\
&=\sum_{n}\frac{\mathrm{d}F_{n}\left(s+\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)}{\mathrm{d}s}\prod_{m\neq n}\Lambda_{m}\left(s+\tau_{m}\,|\,x_{m},\,\boldsymbol{u}_{m}\right).
\end{aligned}$$

Note, that we do not need to take care of truncations since they are independent of s and behave like a constant during differentiation. Reinserting the constant then gives the density of the global survival time

$$\frac{\mathrm{d}F\left(s\,|\,\boldsymbol{\tau},\,\boldsymbol{x}\right)}{\mathrm{d}s}=\sum_{n}\frac{\mathrm{d}F_{n}\left(s+\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)}{\mathrm{d}s}\frac{1}{\Lambda_{n}\left(\tau_{n}\,|\,x_{n},\,\boldsymbol{u}_{n}\right)}\prod_{m\neq n}\frac{\Lambda_{m}\left(s+\tau_{m}\,|\,x_{m},\,\boldsymbol{u}_{m}\right)}{\Lambda_{m}\left(\tau_{m}\,|\,x_{m},\,\boldsymbol{u}_{m}\right)}.$$

A.3.3 Survival Time Parameterization

As we derived an expression of the global survival time in terms of local survival functions $\Lambda_n(s \mid x_n, \boldsymbol{u}_n)$, we are now able to parametrize the local rates of the augmented CTBN. Where in traditional CTBN's, we only have to choose constant rates, in our case, we can assign them arbitrary functions. In this work, choose a desired parametric local survival time distribution and calculate the rates via the known relation

$$\begin{aligned}\frac{\mathrm{d}\Lambda_n(s \mid x_n, \boldsymbol{u}_n)}{\mathrm{d}s} &= -\lambda_n(x_n, s, \boldsymbol{u}_n)\exp\left(-\int_0^s \mathrm{d}\sigma\, \lambda_n(x_n, \sigma, \boldsymbol{u}_n)\right)\\ &= -\lambda_n(x_n, s, \boldsymbol{u}_n)\,\Lambda_n(s \mid x_n, \boldsymbol{u}_n)\end{aligned}$$

$$\lambda_n(x_n, s, \boldsymbol{u}_n) = -\frac{\mathrm{d}\Lambda_n(s \mid x_n, \boldsymbol{u}_n)}{\mathrm{d}s}\frac{1}{\Lambda_n(s \mid x_n, \boldsymbol{u}_n)}.$$

To give an example, we could choose a Weibull distribution giving $\Lambda_n(s \mid x_n, \boldsymbol{u}_n) = bks^{k-1}\exp\left(-bs^k\right)$ with parameters b for the rate and k for the shape. Giving the above equation, we can then calculate $\lambda_n(x_n, s, \boldsymbol{u}_n) = bks^{k-1}$. By assigning individual tuples (b, k) to all combinations n, x_n and $\boldsymbol{u}_n$, we can construct CTBN's with purely Weibull instead of exponential distributions.

A.3.4 Transition Probabilities

Proceeding from here, we obtain the transition probabilities of the global process in terms of local rates, using a similar derivation as for CTSMCs

$$\begin{aligned}&P\left(X_n(t+h) = x_n' \,\middle|\, \bigcup_m X_m(t+h) \neq x_m,\, T_m(t+h) \in [0, 0+h),\, \bigcap_m X_m(t) = x,\, T_m(t) \in [\tau_m, \tau_m + h)\right)\\ &= \frac{P\left(X_n(t+h) = x_n',\, \bigcup_m X_m(t+h) \neq x_m,\, T_m(t+h) \in [0, 0+h) \mid \bigcap_m X_m(t) = x,\, T_m(t) \in [\tau_m, \tau_m + h)\right)}{P\left(\bigcup_m X_m(t+h) \neq x_m,\, T_m(t+h) \in [0, 0+h) \mid \bigcap_m X_m(t) = x,\, T_m(t) \in [\tau_m, \tau_m + h)\right)}\\ &= \frac{P\left(\bigcup_m X_n(t+h) = x_n',\, X_m(t+h) \neq x_m,\, T_m(t+h) \in [0, 0+h) \mid \bigcap_m X_m(t) = x,\, T_m(t) \in [\tau_m, \tau_m + h)\right)}{\sum_{n=1}^N \lambda_n(x_n, \tau_n, \boldsymbol{u}_n)\,h + o(h)} \quad \left(\frac{\text{distributiveness}}{\text{global exit rate}}\right)\\ &= \frac{P\left(X_n(t+h) = x_n',\, T_n(t+h) \in [0, 0+h) \mid \bigcap_m X_m(t) = x,\, T_m(t) \in [\tau_m, \tau_m + h)\right)}{\sum_{n=1}^N \lambda_n(x_n, \tau_n, \boldsymbol{u}_n)\,h + o(h)} \quad \left(\frac{\text{asynchronicity \& absorption}}{}\right)\\ &= \frac{\lambda_n(x_n, \tau_n, \boldsymbol{u}_n; x_n')\,h + o(h)}{\sum_{n=1}^N \lambda_n(x_n, \tau_n, \boldsymbol{u}_n)\,h + o(h)} \quad \left(\frac{\text{local transition rates}}{}\right).\end{aligned}$$

Finally, taking the limit gives us the instantaneous transition probabilities of the global process given the clocks state $\boldsymbol{T}(t) = \boldsymbol{\tau}$

$$\lim_{h \to 0}\left(\frac{\lambda_n(x_n, \tau_n, \boldsymbol{u}_n; x_n')\,h + o(h)}{\sum_{n=1}^N \lambda_n(x_n, \tau_n, \boldsymbol{u}_n)\,h + o(h)}\right) = \frac{\lambda_n(x_n, \tau_n, \boldsymbol{u}_n; x_n')}{\lambda(\boldsymbol{x}, \boldsymbol{\tau})}.$$

A.4 Derivation of the Path Measure

After having obtained the probabilities belonging to the generative process, it is straight forward to give the path density for a single interval. Assume, the associated Markov process of the global process is in state $(\boldsymbol{x}, \boldsymbol{\tau})$ at the beginning of the interval. We keep $\boldsymbol{x}$ for a time exactly s and then, the n-th process transitions from its state x_n to x_n'. The density for this event is then

$$\begin{aligned}p(\boldsymbol{x}', \boldsymbol{\tau}' \mid \boldsymbol{x}, \boldsymbol{\tau}, \boldsymbol{\theta}) &= \frac{\lambda_n(x_n, \tau_n + s, \boldsymbol{u}_n; x_n')}{\sum_{n=1}^N \lambda_n(x_n, \tau_n + s, \boldsymbol{u}_n)}\frac{\mathrm{d}F(s \mid \boldsymbol{\tau}, \boldsymbol{x})}{\mathrm{d}s}\\ &\propto \frac{\lambda_n(x_n, \tau_n + s, \boldsymbol{u}_n; x_n')}{\sum_{n=1}^N \lambda_n(x_n, s + \tau_n, \boldsymbol{u}_n)}\sum_n \lambda_n(x_n, s + \tau_n, \boldsymbol{u}_n)\prod_m \exp\left(-\int_0^{s+\tau_m} \mathrm{d}\sigma\, \lambda_m(x_m, \sigma, \boldsymbol{u}_m)\right)\\ &= \lambda_n(x_n, \tau_n + s, \boldsymbol{u}_n; x_n')\exp\left(-\int_0^{s+\tau_n} \mathrm{d}\sigma\, \lambda_n(x_n, \sigma, \boldsymbol{u}_n)\right)\prod_{m \neq n} \Lambda_m(s + \tau_m \mid x_m, \boldsymbol{u}_m),\end{aligned}$$

with

$$\begin{aligned} \boldsymbol{x}' &= (x_1,\, \ldots,\, x_{n-1},\, x_n',\, x_{n+1},\, \ldots,\, x_N)\,, \\ \boldsymbol{\tau}' &= (\tau_1 + s,\, \ldots,\, \tau_{n-1} + s,\, 0,\, \tau_{n+1} + s,\, \ldots,\, \tau_N + s)\,. \end{aligned} \tag{1}$$

If we further restrict ourselves to processes obeying time-direction independence, we can express the density in terms of transition probabilities $\lambda_n(x_n,\, \tau_n + s,\, \boldsymbol{u}_n;\, x_n') = \theta_n(x_n' \mid x_n,\, \boldsymbol{u}_n)\, \lambda_n(x_n,\, \tau_n + s,\, \boldsymbol{u}_n)$ and obtain after reinserting the constant term representing the truncations

$$\begin{aligned} p(\boldsymbol{x}',\, \boldsymbol{\tau}' \,|\, \boldsymbol{x},\, \boldsymbol{\tau},\, \boldsymbol{\theta}) &= \theta_n(x_n' \mid x_n,\, \boldsymbol{u}_n)\, \lambda_n(x_n,\, \tau_n + s,\, \boldsymbol{u}_n)\, \frac{\exp\left(-\int_0^{s+\tau_n} \mathrm{d}\sigma\, \lambda_n(x_n,\, \sigma,\, \boldsymbol{u}_n)\right)}{\Lambda_n(\tau_n \,|\, x_n,\, \boldsymbol{u}_n)} \prod_{m \neq n} \frac{\Lambda_m(s + \tau_m \,|\, x_m,\, \boldsymbol{u}_m)}{\Lambda_m(\tau_m \,|\, x_m,\, \boldsymbol{u}_m)} \\ &= \theta_n(x_n' \mid x_n,\, \boldsymbol{u}_n)\, \frac{\mathrm{d}F_n(s + \tau_n \,|\, x_n,\, \boldsymbol{u}_n)}{\mathrm{d}s}\, \frac{1}{\Lambda_n(\tau_n \,|\, x_n,\, \boldsymbol{u}_n)} \prod_{m \neq n} \frac{\Lambda_m(s + \tau_m \,|\, x_m,\, \boldsymbol{u}_m)}{\Lambda_m(\tau_m \,|\, x_m,\, \boldsymbol{u}_m)} \end{aligned}$$

We observe, that one part in this expression depends on $\boldsymbol{x}$ and $\boldsymbol{x}'$ and one on $\boldsymbol{x}$, $\boldsymbol{\tau}$ and s. The first is the transition probability associated with the embedded Markov chains of the single processes with time-direction independence. To infer those, we do not need any timing information. On the other hand, to infer the second expression related to the survival times, we do need timing information.

A.5 Likelihood in a Gamma augmented CTBN

In order to derive the likelihood function of a single survival time parameter tuple $(\alpha_n,\, \beta_n)$ associated with the n-th process, its state x and parent state $\boldsymbol{u}$ in a Gamma augmented CTBN, we need to build the product of the respective censored, non-censored and truncated factors of the global likelihood. First consider the sets $\mathbf{X}$ and $\mathbf{T}$ from the main text. The set $\mathbf{X}$ contains samples of the global state at the beginning of each interval. On the other hand, $\mathbf{T}$ contains samples of all clock values at the beginning of each interval. Let $\mathbf{X}_n \subset \mathbf{X}$ and $\mathbf{T}_n \subset \mathbf{T}$ be the subset of samples of states and clock values relevant to the n-th process. This means, that $\mathbf{X}_n$ only contains the sample values from $\mathbf{X}$ corresponding to the n-th process and its parents. The $\mathbf{T}_n$ only contains the sample values of the n-th clock from $\mathbf{T}$. Further, there are no consecutive repetitions in $\mathbf{X}_n$ meaning, that one element in $\mathbf{X}_n$ is not associated with two consecutive intervals. Since we assumed all parameter sets for individual x and $\boldsymbol{u}$ to be independent, samples contained in $\mathbf{X}_n$ and $\mathbf{T}_n$ appear as constants and are normalized out when building the posterior density for $(\alpha_n,\, \beta_n)$. We introduce the following sets: $S_f \equiv \{s_{f,1},\, s_{f,2},\, \ldots\}$ consists of all clock samples $s_{f,m} \in \mathbf{T}_n$ at which the n-th process transitions, S_c consists of all clock samples, where a parent has changed and S_t, which consists of the clock values at the beginning of the intervals. Consider an element $(x,\, \boldsymbol{u}) \in \mathbf{X}_n$ associated with the i-th interval and another element $(x',\, \boldsymbol{u}') \in \mathbf{X}_n$ associated with the $(i+1)$-th interval. If now $x' \neq x$, then the clock sample $s \in \mathbf{T}_n$ associated with the $(i+1)$-th interval is an element of S_f. Otherwise $(x' = x)$, s is an element of S_c. Additionally, $s_t \in \mathbf{t}_n$ associated with the i-th interval is an element of S_t in both cases.

Then, the likelihood is a product of terms of the Gamma p.d.f. $\frac{\beta_n}{\Gamma(\alpha_n)}(\beta_n s_{f,m})^{\alpha_n - 1} \exp(-\beta_n s_{f,m})$ for each $s_{f,m} \in S_f$, of the Gamma survival function $\frac{\Gamma(\alpha_n,\, \beta_n s_{c,m})}{\Gamma(\alpha_n)}$ for each element in S_c, and the reciprocal $\frac{\Gamma(\alpha_n)}{\Gamma(\alpha_n,\, \beta_n s_{t,m})}$ for each element in S_t. Building the product, we then obtain for the posterior update of a full trajectory of the global process associated with the parameters $(\alpha_n,\, \beta_n)$

$$\begin{aligned} &p(\alpha_n,\, \beta_n \mid \mathbf{X}_n,\, \mathbf{T}_n) \\ &\quad \propto \beta_n^{|S_f|} \left(\prod_{m=1}^{|S_f|} \beta_n s_{f,m} \right)^{\alpha_n - 1} \exp\left(-\beta_n \sum_{m=1}^{|S_f|} s_{f,m} \right) \\ &\quad\quad \frac{\Gamma(\alpha_n)^{|S_t|} \prod_{m=1}^{|S_c|} \Gamma(\alpha_n,\, \beta_n s_{c,m})}{\Gamma(\alpha_n)^{|S_f|}\, \Gamma(\alpha_n)^{|S_c|} \prod_{m=1}^{|S_t|} \Gamma(\alpha_n,\, \beta_n s_{t,m})} p(\alpha_n,\, \beta_n) \end{aligned}$$

with an arbitrary prior distribution $p(\alpha_n,\, \beta_n)$. The Weibull posterior update from the main text is constructed in a similar way.

A.6 Inference of Rayleigh augmented CTBN's

Like we have shown, that the prior distribution for a whole CTBN can be given as a product of local prior distributions, the Rayleigh CTBN has a conjugate prior in the form of a product of inverse Gamma distributions under its typical parametrization $\phi_n(x,\, \boldsymbol{u}) \equiv \sigma_n(x,\, \boldsymbol{u})^2$. We can give this by

$$p(\boldsymbol{\phi}) \propto \prod_{n} \prod_{\boldsymbol{u} \in \mathcal{U}_n} \prod_{x \in \mathcal{X}_n} \frac{\beta_n(x,\, \boldsymbol{u})^{\alpha_n(x,\, \boldsymbol{u})}}{\phi_n(x,\, \boldsymbol{u})^{\alpha_n(x,\, \boldsymbol{u})+1}} \exp\left(-\frac{\beta_n(x,\, \boldsymbol{u})}{\phi_n(x,\, \boldsymbol{u})}\right).$$

Because we can perform a normalization after multiplying with the likelihood, we can effectively ignore the constant factor $\beta_n(x,\, \boldsymbol{u})^{\alpha_n(x,\, \boldsymbol{u})}$ and obtain for the posterior update

$$\begin{aligned}
\alpha_n(x,\, \boldsymbol{u}) &\to \alpha_n(x,\, \boldsymbol{u}) + |S_f| \\
\beta_n(x,\, \boldsymbol{u}) &\to \beta_n(x,\, \boldsymbol{u}) + \frac{1}{2}\left(\sum_{m \in S_f \cap S_c} \tau_m^2 - \sum_{m \in S_t} \tau_m^2\right).
\end{aligned}$$

where we can immediately spot the sufficient statistics

$$\begin{aligned}
T_\beta(n,\, x,\, \boldsymbol{u}) &= \frac{1}{2}\left(\sum_{m \in S_f \cap S_c} \tau_m^2 - \sum_{m \in S_t} \tau_m^2\right) \\
T_\alpha(n,\, x,\, \boldsymbol{u}) &= |S_f|.
\end{aligned}$$

Again exploiting the normalization of the resulting expression, by comparison of the product of likelihood and prior with the normalized inverse Gamma distribution, we can formulate the marginal likelihood for structure inference by

$$\int_{\boldsymbol{\Phi}} \mathrm{d}\boldsymbol{\phi}\, p(\mathbf{T} \,|\, \mathbf{X},\, \boldsymbol{\phi})\, p(\boldsymbol{\phi} \,|\, \mathcal{G}) = \prod_{n} \prod_{\boldsymbol{u} \in \mathcal{U}_n} \prod_{x \in \mathcal{X}_n} \frac{\left(\prod_{m \in S_f(n,x,\boldsymbol{u})} \tau_m\right) \beta_n(x,\, \boldsymbol{u})^{\alpha_n(x,\, \boldsymbol{u})}\, \Gamma(\alpha_n(x,\, \boldsymbol{u}) + T_\alpha(n,\, x,\, \boldsymbol{u}))}{\Gamma(\alpha_n(x,\, \boldsymbol{u}))\, (\beta_n(x,\, \boldsymbol{u}) + T_\beta(n,\, x,\, \boldsymbol{u}))^{\alpha_n(x,\, \boldsymbol{u}) + T_\alpha(n,\, x,\, \boldsymbol{u})}}$$

which can be efficiently calculated using Stirling's approximation of the gamma function.

B Configuration of GeneNetWeaver

In this section, we provide the exact settings for GeneNetWeaver from the GRN-scenario. As mentioned in the main text, we chose the maximum time-resolution and simulated until 10^3 units of time. Further, since we perform inference on the latent model directly, we chose an ODE-based simulation without additional noise. We chose "Time series as in DREAM4" and then the following additional settings

Duration of each time series (t_max)	1 000
Number of measured points...	1 001
Perturbations for multifactorial...	Generate new
Noise added after... (measurement error)	None

The resulting time-series' were then preprocessed like mentioned in the main text and the resulting trajectories were then used to train the latent model with a Gamma survival time parametrization.

C Simulation Algorithm

The proposed algorithm for the simulation of the augmented CTBN corresponds to the Gillespie algrorithm for the augmented CTBN and is given below. To draw from a minimum of truncated distributions, we draw from multiple truncated distributions and store the minimum. Additionally, we can draw from the truncated distributions by a simple loop, continuously drawing samples from the original distribution until the outcome is larger than the value of the truncation. In our experiments, no case has occurred, where this lead to significant increases in runtime. However, care must be taken when single processes exhibit an extreme dynamic range. This can potentially necessitate large amounts of draws after unfavorable parent changes.

Algorithm 1: Gillespie algorithm for the augmented CTBN

Result: A sample trajectory of the augmented CTBN from 0 to T

Require θ, ϕ, $\boldsymbol{x}^{(0)}$, $\boldsymbol{\tau}^{(0)}$ (clocks may be all zero);
Set $c \leftarrow 0$;
Set $t^{(0)} \leftarrow 0$;
while $t < T$ **do**
    Draw s from eq. (3) (main text);
    Draw n and x_n' from the categoricals in (5) (main text);
    Set $t^{(c+1)} \leftarrow t^{(c)} + s$;
    if $t^{(c+1)} > T$ **then**
        break;
    else
        Update $\boldsymbol{x}^{(c+1)}$ and $\boldsymbol{\tau}^{(c+1)}$ from $\boldsymbol{x}^{(c)}$ and $\boldsymbol{\tau}^{(c)}$ according to (1) (here in appendix);
        Set $c \leftarrow c + 1$;
    end
    return $\mathbf{X} = \{\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(c)}\}$, $\mathbf{T} = \{\boldsymbol{\tau}^{(0)}, \boldsymbol{\tau}^{(1)}, \ldots, \boldsymbol{\tau}^{(c)}\}$ and $\boldsymbol{t} = \{t^{(0)}, t^{(1)}, \ldots, t^{(c)}\}$;
end

$\mathbf{X}$ then contains the states $\boldsymbol{x}^{(m)}$ in the time-window $\left[t^{(m)}, t^{(m+1)}\right)$. $\mathbf{T}$ contains the clock values at $\boldsymbol{\tau}^{(m)}$ after transition to the new state. Additionally, the value c contains the number of transitions occurred.