Appendix

A common sampling technique often used in this setting is the Metropolis-Hastings algorithm, which is a Markov chain Monte Carlo (MCMC) method. The M-H acceptance probability for moving from state x to state x' is shown below, where each *state* is a DBN.

$$\alpha(x,x') = \min\left\{1, \frac{p(D|x')}{p(D|x)} \times \frac{p(x' \to x)}{p(x \to x')}\right\} = \min\left\{1, \underbrace{\frac{p(D|x')}{p(D|x)}}_{likelihood\ ratio} \times \underbrace{\frac{p(m')p(x|x',m')}{p(m)p(x'|x,m)}}_{proposal\ ratio}\right\}$$

where m is the move type that allows for a transition from state x to x' and m' is the reverse move type for a transition from state x' back to state x. The proposal ratio can be split into two terms: one is the ratio of the proposal probabilities for move types and the other is the ratio of selecting a particular state given the current state and the move type. The choice of scoring metric determines the likelihoods, and often $p(m')$ and $p(m)$ are chosen *a priori* to be simple to calculate or to actually cancel out.

Move type M	Proposal probability	$\frac{p(M')}{p(M)}$	$\frac{p(x|x',M')}{p(x'|x,M)}$
(M_1) add edge to G_1	P_a	$\frac{P_d}{P_a}$	$\frac{(E_1+1)^{-1}}{(np_{max}-E_1)^{-1}} = \frac{np_{max}-E_1}{E_1+1}$
(M_2) delete edge from G_1	P_d	$\frac{P_a}{P_d}$	$\frac{(np_{max}-E_1+1)^{-1}}{E_1^{-1}} = \frac{E_1}{np_{max}-E_1+1}$
(M_3) add edge to Δg_i	P_{ae}	$\frac{P_{de}}{P_{ae}}$	$\frac{m^{-1}(S_i+1)^{-1}}{m^{-1}(S_{max}-S_i)^{-1}} = \frac{S_{max}-S_i}{S_i+1}$
(M_4) delete edge from Δg_i	P_{de}	$\frac{P_{ae}}{P_{de}}$	$\frac{m^{-1}(S_{max}-S_i+1)^{-1}}{m^{-1}S_i^{-1}} = \frac{S_i}{S_{max}-S_i+1}$
(M_5) move edge from Δg_i to Δg_j	P_{me}	1	$\frac{(m-1)^{-1}(\sum_i S_i)^{-1}}{(m-1)^{-1}(\sum_i S_i)^{-1}} = 1$
(M_6) locally shift t_i	P_{st}	1	$\frac{(2d+1)^{-1}}{(2d+1)^{-1}} = 1$
(M_7) merge Δg_i and Δg_{i+1}	P_m	$\frac{P_s}{P_m}$	$\frac{(m-1)^{-1}2(S_i+S_{i+1})^{-1}\binom{S_i+S_{i+1}}{S_i}^{-1}}{(m-1)^{-1}} = \frac{2}{(S_i+S_{i+1})\binom{S_i+S_{i+1}}{S_i}}$
(M_8) split Δg_i	P_s	$\frac{P_m}{P_s}$	$\frac{(m-1)^{-1}}{(m-1)^{-1}(S_i/2)^{-1}\binom{S_i}{x}^{-1}} = (S_i/2)\binom{S_i}{x}$
(M_9) create new Δg_i	P_{ag}	$\frac{P_{dg}}{P_{ag}}$	$\frac{(m+1)^{-1}}{(N-m)^{-1}n^{-2}} = \frac{(N-m)n^2}{m+1}$
(M_{10}) delete Δg_i	P_{dg}	$\frac{P_{ag}}{P_{dg}}$	$\frac{(N-m-1)^{-1}n^{-2}}{m^{-1}} = \frac{m}{(N-m-1)n^2}$

KNKT (for M_1–M_5), KNUT (for M_1–M_6), UNUT (for M_1–M_{10})

Table 1: E_1 is the total number of edges in G_1, S_{max} is the maximum number of transitions allowed in a single transition time, p_{max} is the maximum parent set size, and S_i is the number of edge changes in the set Δg_i. The proposal ratio is the product of the last two columns. The KNKT setting uses moves $(M_1) - (M_5)$, KNUT uses moves $(M_1) - (M_6)$, and UNUT uses moves $(M_1) - (M_{10})$, in each case with the proposal probabilities appropriately normalized to add to 1.

All of the F1-measures for the nsDBNs learned under the UNUT setting are shown in Table 2 below.

A

λ_s

1	2	3	4	5	λ_m
0.4341	0.9423	0.9469	0.9738	**0.9912**	1
0.6760	0.9562	0.9553	0.9906	0.9909	2
0.9206	0.9553	0.9729	0.9731	0.9905	5
0.9264	0.9550	0.9657	0.9829	0.9791	10
0.8804	0.8806	0.9042	0.8922	0.8807	50

B

λ_s

1	3	5	λ_m
0.9489	0.9510	0.9468	1
0.9377	0.9521	0.9356	2
0.9531	0.9459	0.9398	5

Table 2: F1-measure for different values of λ_s and λ_m under the UNUT setting. F1-measures over 0.9 are shaded in light gray and the best score is shown in bold. **A:** F1-measures for the nine variable dataset defined by the nsDBN in Figure 1A. **B:** F1-measures for the large 100 variable dataset.