Macroscopic Models of Clique Tree Growth for Bayesian Networks

Ole J. Mengshoel
RIACS
NASA Ames Research Center
Mail Stop 269-3
Moffett Field, CA 94035
omengshoel@riacs.edu

Abstract

In clique tree clustering, inference consists of propagation in a clique tree compiled from a Bayesian network. In this paper, we develop an analytical approach to characterizing clique tree growth as a function of increasing Bayesian network connectedness, specifically: (i) the expected number of moral edges in their moral graphs or (ii) the ratio of the number of non-root nodes to the number of root nodes. In experiments, we systematically increase the connectivity of bipartite Bayesian networks, and find that clique tree size growth is well-approximated by Gompertz growth curves. This research improves the understanding of the scaling behavior of clique tree clustering, provides a foundation for benchmarking and developing improved BN inference algorithms, and presents an aid for analytical trade-off studies of tree clustering using growth curves.

Introduction

Bayesian networks play a central role in a wide range of automated reasoning applications, including in diagnosis, probabilistic risk analysis, information fusion, error correction coding, and object recognition. Clique tree clustering, where inference takes the form of propagation in a clique tree compiled from a Bayesian network (BN), is currently among the most prominent Bayesian network inference algorithms (Lauritzen & Spiegelhalter 1988; Andersen *et al.* 1989; Shenoy 1989). The performance of tree clustering algorithms depends on a BN's treewidth or the optimal maximal clique size of a BN's induced clique tree (Dechter & Pearl 1987; Darwiche 2001; Dechter & Fattah 2001). The performance of other exact BN inference algorithms also depends on treewidth.

A key research question is, then, how the clique tree size of a BN (and consequently, inference time) depends on some measure of the BN's connectedness. One way to investigate this is through the use of distributions of problem instances (Suermondt & Cooper 1990; Ide, Cozman, & Ramos 2004; Mengshoel, Wilkins, & Roth 2006). Taking this approach, and varying the ratio C/V between the number of clauses C and the number of variables V in CNF formulas, an easy-hard-easy pattern was established for the Davis-Putnam search algorithm (Mitchell, Selman, & Levesque 1992). Using a similar approach in a BN setting, an easy-hard-harder pattern has been observed for clique tree clustering (Mengshoel, Wilkins, & Roth 2006).

In this paper, we develop macroscopic models of clique tree growth by means of restricted growth curves, and illustrate our approach using bipartite BNs. A key finding is that Gompertz growth curves are justified on theoretical grounds and also fit very well to experimental data generated using the BPART algorithm (Mengshoel, Wilkins, & Roth 2006). Our analysis using Gompertz growth curves is novel; they are common in biological and medical research (Banks 1994; Lindsey 2004) but have not previously been used to characterize clique tree growth. We provide improved analysis compared to previous research, where an easy-hard-harder pattern and approximately exponential growth as a function of C/V-ratio were established (Mengshoel, Wilkins, & Roth 2006).

This research is significant for the following reasons. First, analytical growth curves improve the understanding of clique tree clustering's performance. Consider Kepler's three laws of planetary motion, developed using Brahe's observational data of planetary movement. There is a need to develop similar laws for clique tree clustering's performance, and in this paper we obtain laws in the form of Gompertz growth curves for certain bipartite BNs (Mengshoel, Wilkins, & Roth 2006). These curves give significantly better fit to the raw data than previously employed exponential curves, provide better insight into the underlying mechanisms of the algorithm, and may be used to approximately predict the performance of clique tree clustering. Second, growth curves can be used to summarize performance of different BN inference algorithms or different implementations of the same algorithm on benchmarks, and thereby aid in evaluations. Suppose that the growth curves $g(x)$ and $h(x)$ were obtained by benchmarking slightly different clique tree algorithms. Compared to using only raw data, it may be easier to understand the performance difference between the two algorithms by comparing the parameter values of $g(x)$ and $h(x)$. Third, growth curves provide estimates of resource consumption in terms of clique tree size. Resource bounds, for example on memory size and inference time, represent requirements from applications and can also be expressed in terms of clique tree size. Hence,

this approach enables trade-off studies of resource consumption versus resource bounds, which is important in resource-bounded reasoners (Mengshoel 2007).

The rest of this paper is organized as follows. After first introducing notation and concepts, we study BNs in which the number of moral edges and clique tree sizes are characterized by random variables. Second, we describe how growth curves can provide a macroscopic model of how clique trees grow as a function of C/V-ratio or expected number of moral edges. Third, we present experiments with varying number of BN root and leaf nodes. We compare different mathematical models of growth, and find that Gompertz growth curves give the best fit to experimental results. Finally, we conclude and suggest future research directions.

Background

A Bayesian network (BN) is a tuple $\beta = (\boldsymbol{X}, \boldsymbol{E}, \boldsymbol{P})$, where $(\boldsymbol{X}, \boldsymbol{E})$ is a DAG with an associated set of conditional probability distributions $\boldsymbol{P} = \{\Pr(X_1 \mid \Pi_{X_1}), \ldots, \Pr(X_n \mid \Pi_{X_n})\}$. Here, $\Pr(X_i \mid \Pi_{X_i})$ is the conditional probability distribution for $X_i \in \boldsymbol{X}$. Let π_{X_i} represent the instantiation of the parents Π_{X_i} of X_i. The independence assumptions encoded in $(\boldsymbol{X}, \boldsymbol{E})$ imply the joint probability distribution

$$\Pr(\boldsymbol{x}) = \prod_{i=1}^{n} \Pr(x_i \mid \pi_{X_i}), \qquad (1)$$

where $\Pr(\boldsymbol{x}) = \Pr(X_1 = x_1, \ldots, X_n = x_n)$.

A clique tree β''' is constructed from a BN in the following way by the HUGIN clique tree clustering algorithm (Lauritzen & Spiegelhalter 1988; Andersen *et al.* 1989). First, an initial *moral graph* β' is constructed by making an undirected copy of β and then augmenting it with moral edges as follows. For each node $X \in \boldsymbol{X}$, HUGIN adds to β' a *moral edge* between each pair of nodes in Π_X if no such edge already exists in β'. Second, HUGIN creates a triangulated graph β'' by heuristically adding fill-in edges to β' such that no chordless cycle of length greater than three exists. Third, a clique tree β''' is created from the triangulated graph β''. For any two nodes F and H in the clique tree, all nodes between them contain $F \cap H$. Using β''', HUGIN can perform *belief updating* (compute marginals) (Lauritzen & Spiegelhalter 1988) or *belief revision* (MPE computation), and the compilation and propagation times are in both cases strongly dependent upon the size of β'''. The complexity of many exact BN inference algorithms — including tree clustering algorithms, conditioning algorithms, and elimination algorithms — has been found to depend on treewidth ϖ^* or on minimal maximal clique size h^*, where $\varpi^* = h^* - 1$ (Lauritzen & Spiegelhalter 1988; Dechter & Fattah 2001). Treewidth computation is NP-complete (Arnborg, Corneil, & Proskurowski 1987), and greedy triangulation heuristics that compute upper bounds on treewidth are typically used in practice (Koster, Bodlaender, & van Hoesel 2001).

Using randomly generated problem instances, one can systematically investigate BN inference algorithms (Suermondt & Cooper 1990; Ide, Cozman, & Ramos 2004;
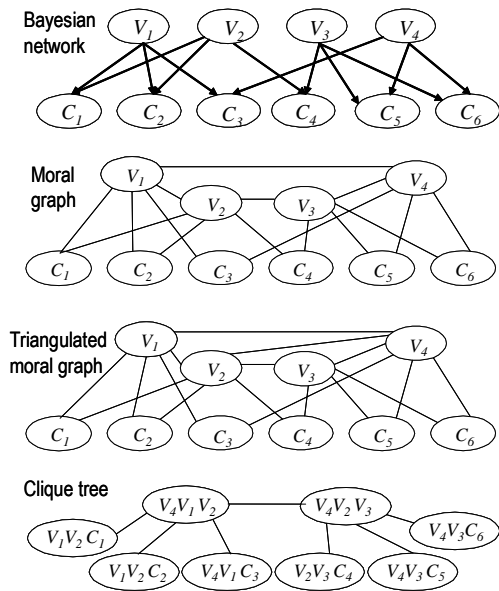

Figure 1: Compilation of BPART BN β (top) to clique tree β''' (bottom). There is a loop (V_1, V_2, V_3, V_4) in the moral graph β', leading to a fill-in edge (V_2, V_4) in triangulated graph β'', which again leads to cliques $\{V_4, V_1, V_2\}$ and $\{V_4, V_2, V_3\}$ in the clique tree β'''.

Mengshoel, Wilkins, & Roth 2006). While our approach is general, we study bipartite BNs in detail. These are BNs in which $\boldsymbol{X}$ is partitioned into root nodes $\boldsymbol{V}$ and leaf nodes $\boldsymbol{C}$; all edges $\boldsymbol{E}$ are from nodes in $\boldsymbol{V}$ to nodes in $\boldsymbol{C}$. These BNs are generated randomly using the BPART algorithm (Mengshoel, Wilkins, & Roth 2006), which is a generalization of an approach to randomly generate problem instances for satisfiability (Mitchell, Selman, & Levesque 1992). Here is an example of a small BPART BN.

Example 1 (BPART Class A BN) *Figure 1 shows how a* BPART *BN may be compiled into a clique tree. For each BN leaf node $C \in \{C_1, C_2, C_3, C_4, C_5, C_6\}$, a clique is created. In addition, there are two cliques containing BN root nodes only, namely the cliques $\{V_1, V_2, V_4\}$ and $\{V_2, V_3, V_4\}$.*

This bipartite BN illustrates the crucial formation of cycles in a BN's moral graph and the resulting generation of fill-in edges. In larger BNs, it is important but also very difficult to understand and predict clique tree clustering's cycle-generation and fill-in processes.

Bipartite BNs are found in applications, including in medical diagnosis BNs (Shwe *et al.* 1991) and in error correction coding (MacKay 2002). In addition, general BNs often have non-trivial bipartite components. The BPART algorithm operates as follows. First, $V = |\boldsymbol{V}|$ root nodes and $C = |\boldsymbol{C}|$ leaf nodes, all with S states, are created. For each leaf node, P parent nodes $\{X_1, \ldots, X_P\}$ are picked uniformly at random without replacement among the V root nodes. Tree clustering's moralization step in turn ensures that there are edges between all P root nodes that share a

leaf node. To keep the discussion succinct, we often say that BPART picks moral edges and omit explicit mention of tree clustering's moralization step. Conditional probability tables (CPTs) of all nodes are also constructed by BPART; however in this paper we focus on the impact of the structural parameters V, C, P, and S on clique tree size. Consequently, the signature BPART(V, C, P, S) is used. The total number of BN nodes is $N = C + V$.

From Bayesian Networks to Clique Trees

The balls and bins model, where balls are placed uniformly at random into bins, turns out to be useful in our analysis of tree clustering's moralization step. Let m denote the number of balls and n denote the number of bins. Further, let Y be a random variable representing the number of empty and occupied bins respectively. The expected number of occupied bins Y is

$$E(Y) = n \left(1 - (1 - 1/n)^m\right). \qquad (2)$$

In our analysis of clique tree clustering, bins are all possible edges in the moral graph and balls are non-root nodes that induce actual edges in the moral graph.

Definition 2 (Edge-bin) *Let $\mathbf{V}$ be the non-leaf nodes in a BN. An edge-bin is an edge in the moral graph, induced by two non-leaf nodes $\{V_1, V_2\}$, where $V_1, V_2 \in \mathbf{V}$ and $V_1 \neq V_2$. The set of all edge-bins is $\{\{V_1, V_2\} \mid V_1, V_2 \in \mathbf{V}\}$.*

Definition 3 (Edge-ball) *Let $\mathbf{V}$ be the non-leaf nodes in a BN. An edge-ball is the set of moral edges induced by the P parents $\Pi_C = \{V_1, ..., V_P\} \subseteq \mathbf{V}$ of a non-root node C: $\{\{V_1, V_2\}, \{V_1, V_3\}, ..., \{V_{P-1}, V_P\}\}$.*

For bipartite BNs, the non-leaf nodes are the root nodes and the non-root nodes are the leaf nodes. In the BPART(V, C, 2, S) model, where $P = 2$, all edge-bins are uniformly and repeatedly eligible for placing edge-balls into. In other words, we have sampling with replacement. Intuitively, as the C/V-ratio or other measures of connectivity increase (Mitchell, Selman, & Levesque 1992; Mengshoel, Wilkins, & Roth 2006), it gets more and more likely that the same moral edges get picked two or more times. This intuitive argument is formalized in this theorem.

Theorem 4 (Moral edges, two parents) *Consider BNs generated using* BPART*(V, C, P, S). Let the number of moral edges created by tree clustering be a random variable W. For $P = 2$ we have:*

$$E(W) = \binom{V}{2} \left(1 - \left(1 - 1 \left/ \binom{V}{2}\right.\right)^{C}\right). \qquad (3)$$

Proof. We use the balls and bins model and let the edge-balls be the $m = C$ leaf nodes. The edge-bins are all possible $n = \binom{V}{2}$ moral edges in a bipartite graph with V root nodes. Plugging m and n into (2) gives (3). ∎

In contrast to the C/V-ratio (Mitchell, Selman, & Levesque 1992; Mengshoel, Wilkins, & Roth 2006), the expectation $E(W)$ in (3) takes into account how tree clustering

picks parents among pairs of BN root nodes with replacement. For low values of C/V or E/V, the effect of replacement is minor, but for large C/V or E/V the difference may be substantial as illustrated in the following examples.

Example 5 ($C = 30$ leaf nodes) *Let $V = 30$, $C = 30$, and $P = 2$. The expected number of moral edges is $E(W) = 28.99$ using (3).*

Example 6 ($C = 300$ leaf nodes) *Let $V = 30$, $C = 300$, and $P = 2$. The expected number of moral edges is $E(W) = 216.91$ using (3).*

The difference between these two examples is important, because the number of moral edges as well as how they are connected and form cycles (which again need to be triangulated) are key drivers of clique tree size.

For $P > 2$, BPART uses a combination of sampling with replacement and sampling without replacement. In order to simplify analysis, we now introduce a variant BPART^{+} which works exactly as BPART except that the P parent nodes are picked with replacement.

Theorem 7 (Moral edges, any number of parents)
Consider BNs generated using BPART^{+}*(C, V, P, S), and let the number of moral edges created by tree clustering be a random variable Z. Then:*

$$E(Z) = \binom{V}{2} \left(1 - \left(1 - 1 \left/ \binom{V}{2}\right.\right)^{C\binom{P}{2}}\right). \qquad (4)$$

Proof. We use the balls and bins model, and again the number of edge-bins is $n = \binom{V}{2}$ in a bipartite graph with V root nodes. Since BPART^{+} employs sampling with replacement, the number of edge-balls is $m = C \times \binom{P}{2}$. Plugging m and n into (2) gives (4). ∎

Theorem 7 is clearly a generalization of Theorem 4. Further, $E(Z)$ can be used as an approximation for $E(W)$ for BPART(V, C, P) for $P > 2$, since it is well-known that sampling with replacement approaches sampling without replacement as the number of objects sampled from (here, the V root nodes) tends to infinity.

Why are the above balls and bins models of BN moralization interesting? The reason is that we are concerned with the possible causes, at a macroscopic level, of inference difficulty for tree clustering, and the expected number of moral edges is one such cause. When it comes to the effect, namely tree clustering performance, it is natural to minimize the size of the maximal clique. Since this is hard (Arnborg, Corneil, & Proskurowski 1987), current algorithms including HUGIN use heuristics that upper bound optimal maximal clique size and clique tree size. Such upper bounds on clique tree size are just referred to as clique tree sizes in the following.

For bipartite BNs, including BPART BNs, there are two types of nodes in the clique tree as reflected in the following definition.[1]

---

[1]This and the following definition can easily be generalized to cover arbitrary BNs, but in order to be consistent with the experiments we use bipartite BNs in these definitions.

Definition 8 (Root clique, mixed clique) *Consider   a clique tree β''' constructed from a bipartite BN β. A clique in β''' is denoted a* root clique *if all the BN nodes in that clique are root nodes in β. A clique in β''' is denoted a* mixed clique *if the BN nodes in that clique are both root nodes and leaf nodes in β.*

Random variables K_T, K_R, and K_M are used to represent the total clique tree size, the size of all root cliques, and the size of all mixed cliques respectively:

$$K_T = K_R + K_M. \qquad (5)$$

Total clique tree size is the sum of the clique sizes, as is appropriate for HUGIN. We use (5) and linearity of expectation to obtain

$$E(K_T) = E(K_R) + E(K_M)$$
$$\mu_T = \mu_R + \mu_M. \qquad (6)$$

When varying one or more of BPART's parameters we sometimes make that explicit in (6). For instance, the notation $\mu_R(C)$ or $\mu_M(C)$ means that C is varied while V, P, and S are kept constant. In the experimental part of this paper, μ_R will be estimated using its sample mean $\hat{\mu}_R$. Collections of such sample means are then used to construct growth curves.

Clique Tree Growth Curves

Here, we develop macroscopic and restricted models of clique tree growth that extend exponential growth curves (Mengshoel, Wilkins, & Roth 2006) that model unrestricted growth. Even though Bayesian networks and clique trees are discrete structures, we use continuous growth models in order to facilitate analysis.

Definition 9 (Clique tree growth curve) *Let $g_R(x)$ be the growth curve for all root cliques and $g_M(x)$ the growth curve for all mixed cliques. The (total) clique tree growth curve for a bipartite BN is defined as*

$$g_T(x) = g_R(x) + g_M(x).$$

A number of sigmoidal growth curves ("S-curves") have been used to model restricted growth, including the logistic, Gompertz, Complementary Gompertz, and Richards growth curves (Banks 1994; Lindsey 2004). It turns out that the Gompertz growth curve gives a good approximation for clique tree growth.

Definition 10 (Gompertz growth curve) *Let ζ , $\gamma \in \mathbb{R}$ with $\zeta > 0$ and $\gamma > 0$. The Gompertz growth curve is*

$$g(x) = g(\infty)e^{-\zeta e^{-\gamma x}}, \qquad (7)$$

where $g(\infty)$ is the asymptote as $x \to \infty$.

The derivative $g'(x)$ of the Gompertz growth curve

$$g'(x) = \frac{d}{dx}g(x) = g(\infty)\zeta\gamma e^{-\gamma x}e^{-\zeta e^{-\gamma x}},$$

is an expression of the *growth rate* of $g(x)$; clearly $g'(x) > 0$ given our assumptions in Definition 10.
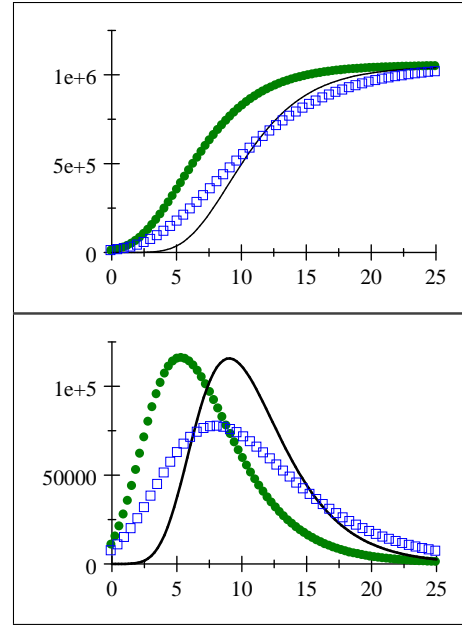

Figure 2: *Top:* Gompertz curves $g_1(x) = 2^{20}e^{-5e^{-0.3x}}$ (green dotted curve), $g_2(x) = 2^{20}e^{-15e^{-0.3x}}$ (black solid curve), and $g_3(x) = 2^{20}e^{-5e^{-0.2x}}$ (blue boxed curve). *Bottom:* Growth rates $g_1'(x)$, $g_2'(x)$ and $g_3'(x)$ for the Gompertz growth curves.

In Figure 2 we investigate how the parameters $g(\infty)$, ζ, and γ impact the shape of Gompertz curves. The factor $g(\infty) = 2^{20}$ is obtained, for example, by considering bipartite BNs with $V = 20$ binary ($S = 2$) root nodes. Figure 2 also shows how the growth rate $g'(x)$ changes when the parameters ζ and γ are varied. Let us first vary ζ as shown in Figure 2. By increasing ζ from $\zeta = 5$ to $\zeta = 15$ while keeping $\gamma = 0.3$ constant, the x-location of maximal growth rate $g'(x)$ is increased as well. However, the value of $g'(x)$ at its maximum does not change. Let us next vary γ as illustrated in Figure 2. As γ decreases from $\gamma = 0.3$ to $\gamma = 0.2$, while $\zeta = 5$ is kept constanc, the x-location of maximal $g'(x)$ increases. In addition, the maximal value of $g'(x)$ decreases with γ decreasing, and generally growth gets more gradual as γ decreases.

In the context of BNs, the independent variable x for the growth curve $g(x)$ may be parametrized using $x = C$, $x = C/V$, $x = E/V = CP/V$, or $x = E(W)$, depending on the data available and the purpose of the model. We now introduce a Gompertz growth curve for BPART.

Theorem 11 (BPART Gompertz growth curve) *The total growth curve $g_T(x)$ for* BPART*(V, C, P, S), assuming Gompertz growth for root cliques and where $x = C$ is the independent variable, is*

$$g_T(x) = S^V e^{-\zeta e^{-\gamma x}} + xS^{P+1}. \qquad (8)$$

Proof. Since BPART BNs are bipartite, the growth curve has the form $g_T(x) = g_R(x) + g_M(x)$, where $g_R(x) =$

$g_R(\infty)e^{-\zeta e^{-\gamma x}}$ since we have the Gompertz growth curve. For BPART(V, C, P, S) we have $g_R(\infty) = S^V$, and therefore $g_R(x) = S^V e^{-\zeta e^{-\gamma x}}$ for appropriate choices of ζ and γ. Total mixed clique size is $C \times S^{P+1}$ (Mengshoel, Wilkins, & Roth 2006), and hence $g_M(x) = xS^{P+1}$. By forming $g_R(x) + g_M(x)$ we obtain the desired result (8). ∎

Analytical growth models or growth curves have been used to model organisms in biology and medicine, growth of technology use or penetration, and growth of organizations or societies including the Web (Banks 1994; Lindsey 2004). However, our use of growth curves to model how clique tree size grows with $x = C$, $x = C/V$, or $x = E(W)$ is, to our knowledge, novel.

The Gompertz growth curve can be derived by solving the differential equation $dg(x)/dx = ag(x)$, where a is a growth coefficient (Banks 1994). Here, a is not constant but exponentially decreasing, formally $da/dx = -ka$ for $k > 0$. These two equations can be solved to obtain (7); see (Banks 1994). While a detailed study is beyond the scope of this paper, it appears plausible that these differential equations reflect, at a macroscopic level, clique tree clustering's formation of cycles in a moral graph β' along with the generation of fill-in edges. Once one cycle appears in β', there may be many cycles appearing, all needing fill-in edges. Thus, once cycle formation starts in β', a faster than exponential growth in root clique tree size $g_R(x)$ is realistic and indeed supported by previous experimental results (Mengshoel, Wilkins, & Roth 2006). This rapid growth can be captured by Gompertz growth curves.

We emphasize that Gompertz curves do not always provide accurate models of clique tree growth. In particular, the assumption $g'_R(x) > 0$ is not valid for very small $x = C$. Consider the first few BN leaf nodes added by BPART. When there is no leaf node and $x = 0$, clearly $\mu_R(0) = V$ and $\mu_M(0) = 0$. When there is one leaf node with P parents and $x = 1$, $\mu_R(1) = V - P$ and $\mu_M(1) = S^{P+1}$. Since $\mu_R(0) > \mu_R(1)$, the contribution of the root cliques to the total clique tree size in fact decreases from $x = 0$ to $x = 1$, and clearly this is not consistent with $g'_R(x) > 0$. However, this early stage of growth is perhaps the least interesting since the total clique tree size is small and not a concern in applications. Consequently, we consider this a minor limitation and use $C/V \geq 1/2$ in our experiments below.

Finally, we note that the Gompertz growth curve has a linear form, defined as follow (Lindsey 2004).

Definition 12 (Gompertz linear form) *The Gompertz linear form is*

$$\ln\left(-\ln\frac{g(x)}{g(\infty)}\right) = \ln(\zeta) - \gamma x \qquad (9)$$

Using (9), the Gompertz curve parameters ζ and γ in (7) can be estimated from data using linear regression, as we will see in the next section.

Experiments

In the experiments we address the following questions: How well do Gompertz growth curves match sample data in the form of clique trees generated from bipartite Bayesian networks using tree clustering? How well do Gompertz growth curves fit sample data compared to alternative growth curve models? In answering these questions, we extend previous research (Mengshoel, Wilkins, & Roth 2006) and use a greater range of values for C/V, investigate $V = 20$ and $V = 30$, and introduce $E(W)$ in addition to C/V as the independent parameter. Clique trees were generated using an implementation of the HUGIN clique tree clustering algorithm. Clique trees were optimized heuristically, using the minimum fill-in weight triangulation heuristic, as treewidth computation is NP-complete.

Comparison between Growth Models

The purpose of the first set of experiments was to compare the Gompertz growth model with a few alternatives: Exponential, logistic, and complementary Gompertz. Here, we report on Bayesian networks generated using the signature BPART(30, C, 2, 2) with varying values for C. For each C/V-level, 100 BNs were sampled using BPART.

We now present the results of the HUGIN experiments. At the top of Figure 3, sample means $\hat{\mu}_R$ and corresponding points from analytical growth curves as a function of $E(W)$ are presented. The bottom of Figure 3 shows how the growth curves at the top were obtained using linear forms such as (9). The following Gompertz growth curve was ob-
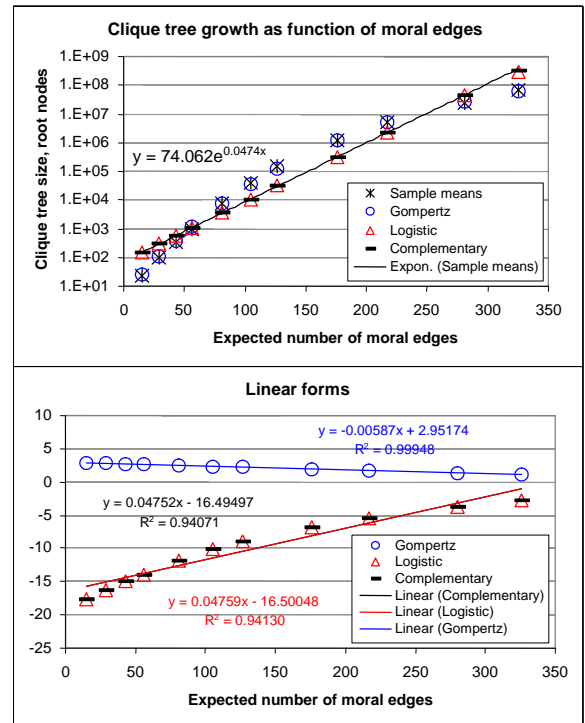

Figure 3: Experimental results for bipartite BNs with $V = 30$ root nodes and varying number of leaf nodes. *Top*: Comparison of Gompertz and other growth curves with the sample means. *Bottom*: Linear forms showing how the growth curves were obtained.

tained

$$g_R(x) = 2^{30} \times \exp(-19.\,14 \times \exp(-0.005874x)),$$

where $x = E(W)$. The parameters ζ and γ were for the other growth curves computed in a similar manner. Clearly, the Gompertz curve fits the data much better than the alternative growth curves analyzed, with $R^2 = 0.9995$ versus $R^2 = 0.9413$ (for logistic) and $R^2 = 0.9407$ (for complementary Gompertz).

Gompertz Growth Model Details

In a second set of experiments, Bayesian networks were generated using BPART(20, C, 2, 2) with varying values for C. For each C/V-level, 100 BNs were sampled using BPART. Using this relatively low value for V allowed us to generate BNs for which the generated clique trees did not exhaust the computer's memory even for very large C, thus supporting a comprehensive analysis using Gompertz growth curves with both $x = C/V$ and $x = E(W)$ as independent variables.

Figure 4 illustrates the results of these HUGIN experiments. In the top row of Figure 4, sample means and corresponding points from the Gompertz growth curve

$$g_R(x) = 2^{20} \times \exp(-9.\,906 \times \exp(-0.1118x)),$$

as a function of $x = C/V$ are presented along with an exponential regression curve. The values of $\zeta = e^{2.293} = 9.\,906$ and $\gamma = 0.1118$ were obtained from the Gompertz linear form as illustrated to the top right in Figure 4, based on sample means for the clique tree root cliques and the linear regression result $\ln(\zeta) - \gamma x = -0.1118x + 2.293$.

In the bottom row of Figure 4, we plot the expected number of moral edges $E(W)$ along the x-axis. Note that the right-most sample average in the bottom row of Figure 4, at $x = E(W) \approx 123$, corresponds to the sample average at $C/V = 10$ in the top row of Figure 4. We present sample means along with the corresponding points from a Gompertz growth curve as a function of $E(W)$; an exponential regression curve is presented as a baseline. Here, the Gompertz growth curve was empirically determined to be

$$g_R(x) = 2^{20} \times \exp(-12.\,43 \times \exp(-0.01187x)),$$

where $x = E(W)$. The parameters ζ and γ were computed in a similar manner to above and as summarized to the bottom right in Figure 4.

Figure 4 clearly shows the improved fit provided by Gompertz curves compared to exponential curves. Further, $x = E(W)$ provides a better fit than $x = C/V$ but for a narrower domain. As a heuristic, one can say that $x = E(W)$ is preferable for local growth models for small values of x, while $x = C/V$ is better for global models and for large x.

Conclusion and Future Work

Much progress has recently been made, both in the area of Bayesian network (BN) reasoning algorithms and in the area of applications of BNs. However, a precise understanding of how varying structural parameters cause BNs to have large treewidths or optimal clique tree sizes is still lagging. In this paper, we have investigated the tree clustering approach to BN inference by employing a growth curve approach. We have characterized the growth of clique tree size as a function of an increase in the expected number of moral edges or an increasing C/V-ratio, for example resulting from an increase in the number of leaf nodes in bipartite BNs. Gompertz growth curves have, for the bipartite BNs investigated, been shown to give excellent fit to empirical clique tree data and they appear theoretically plausible as well.

Areas for future work include the following. First, this type of approach may be utilized in trade-off studies and in knowledge-based model construction, where in both cases there is uncertainty regarding the structure of the BNs processed. Second, it would be interesting to develop more comprehensive analytical models including models for larger classes of BNs.

References

Andersen, S. K.; Olesen, K. G.; Jensen, F. V.; and Jensen, F. 1989. HUGIN—a shell for building Bayesian belief universes for expert systems. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, volume 2, 1080–1085.

Arnborg, S.; Corneil, D. G.; and Proskurowski, A. 1987. Complexity of finding embeddings in a k-tree. *SIAM Journal of Algebraic and Discrete Meththods* 8:277–284.

Banks, R. B. 1994. *Growth and Diffusion Phenomena*. New York: Springer.

Darwiche, A. 2001. Recursive conditioning. *Artificial Intelligence* 126(1-2):5–41.

Dechter, R., and Fattah, Y. E. 2001. Topological parameters for time-space tradeoff. *Artificial Intelligence* 125(1-2):93–118.

Dechter, R., and Pearl, J. 1987. Network-based heuristics for constraint satisfaction problems. *Artificial Intelligence* 34(1):1–38.

Ide, J. S.; Cozman, F. G.; and Ramos, F. T. 2004. Generating random Bayesian networks with constraints on induced width. In *Proceedings of the 16th European Conference on Artificial Intelligence*, 323–327.

Koster, A. M. C. A.; Bodlaender, H. L.; and van Hoesel, S. P. M. 2001. Treewidth: Computational experiments. In Broersma, H.; Faigle, U.; Hurink, J.; and Pickl, S., eds., *Electronic Notes in Discrete Mathematics*, volume 8. Elsevier Science Publishers.

Lauritzen, S., and Spiegelhalter, D. J. 1988. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society series B* 50(2):157–224.

Lindsey, J. K. 2004. *Statistical Analysis of Stochastic Processes in Time*. Cambridge: Cambridge.
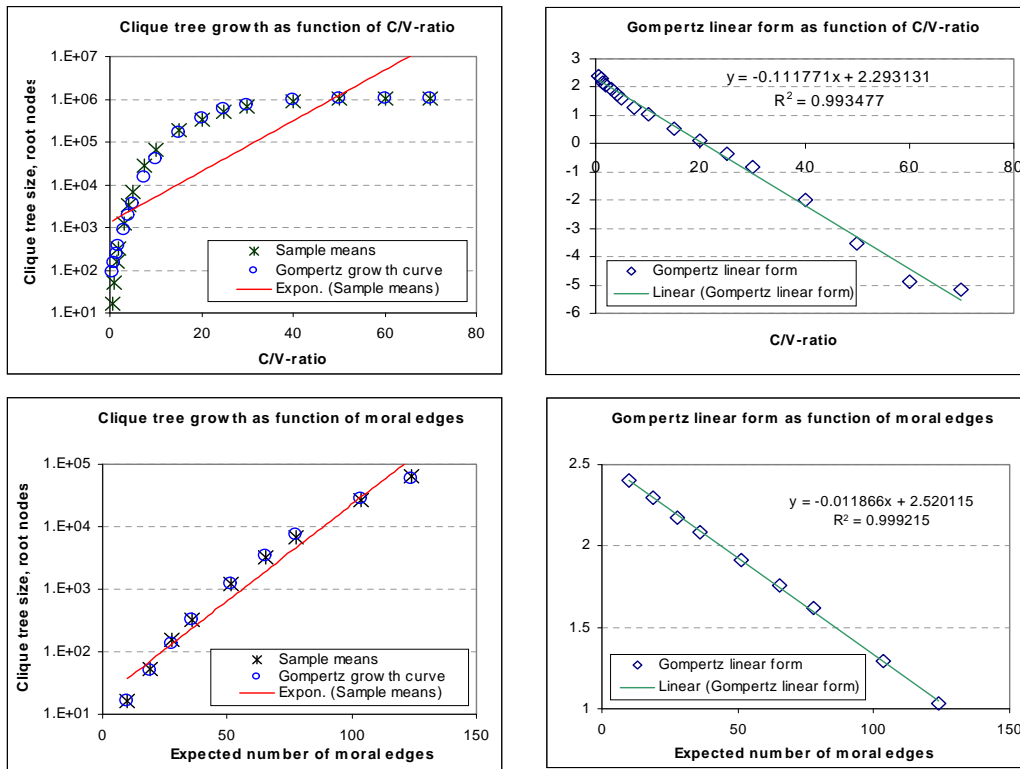
MacKay, D. J. C. 2002. *Information Theory, Inference and*

Figure 4: Empirical results for bipartite Bayesian networks generated with $V = 20$ root nodes and a varying number of leaf nodes C. *Top left:* Gompertz growth curve as a function of the C/V-ratio. *Top right:* Gompertz growth curve's linear form as a function of the C/V-ratio; used to create the Gompertz growth curve to the left. *Bottom left:* Gompertz growth curve as a function of $E(W)$. *Bottom right:* Gompertz growth curve's linear form as a function of $E(W)$; used to create the Gompertz growth curve to the left.

Learning Algorithms. Cambridge, UK: Cambridge University Press.

Mengshoel, O. J.; Wilkins, D. C.; and Roth, D. 2006. Controlled generation of hard and easy Bayesian networks: Impact on maximal clique tree in tree clustering. *Artificial Intelligence* 170(16-17):1137–1174.

Mengshoel, O. J. 2007. Designing resource-bounded reasoners using Bayesian networks: System health monitoring and diagnosis. In *Proceedings of the 18th International Workshop on Principles of Diagnosis (DX-07)*.

Mitchell, D.; Selman, B.; and Levesque, H. J. 1992. Hard and easy distributions of SAT problems. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, 459–465.

Shenoy, P. P. 1989. A valuation-based language for expert systems. *International Journal of Approximate Reasoning* 5(3):383–411.

Shwe, M.; Middleton, B.; Heckerman, D.; Henrion, M.; Horvitz, E.; Lehmann, H.; and Cooper, G. 1991. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: I. The probabilistic model and inference algorithms. *Methods of Information in Medicine* 30(4):241–255.

Suermondt, H. J., and Cooper, G. F. 1990. Probabilistic inference in multiply connected belief networks using loop cutsets. *International Journal of Approximate Reasoning* 4:283–306.