Random Bayesian networks with bounded indegree

Eunice Yuh-Jie Chen
Cognitive Systems Laboratory
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
eyjchen@cs.ucla.edu

Judea Pearl
Cognitive Systems Laboratory
Department of Computer Science
University of California, Los Angeles
Los Angeles, CA 90095
judea@cs.ucla.edu

Abstract

Bayesian networks (BN) are an extensively used graphical model for representing a probability distribution in artificial intelligence, data mining, and machine learning. In this paper, we propose a simple model for large random BNs with bounded indegree, that is, large directed acyclic graphs (DAG) where the edges appear at random and each node has at most a given number of parents. Using this model, we can study useful asymptotic properties of large BNs and BN algorithms with basic combinatorics tools. We estimate the expected size of a BN, the expected size increase of moralization, the expected size of the Markov blanket, and the maximum size of a minimal d-separator. We also provide an upper bound on the average time complexity of an algorithm for finding a minimal d-separator. In addition, the estimates are evaluated against BNs learned from real world data.

1 INTRODUCTION

A Bayesian network (BN) is a directed acyclic graph (DAG) that provides a concise representation of a probability distribution. It is used extensively in artificial intelligence, data mining, and machine learning. Because of the increasing amount of data now available, today many BNs contain a large number of nodes. Though graphical models have been proposed for large BNs with repeated structures, such as the dynamic BN and plate model, a wide range of important applications nonetheless involve large BNs with no repeated structures (Cohen and Havlin, 2010; Guo and Hsu, 2002; Jones et al., 2005; Koller and Friedman, 2009; Schadt et al., 2010).

In this paper, we propose a simple model for the distribution of large random BNs, that is, DAGs where the edges appear at random, each node has at most k parents, and k is much less than the number of nodes n. Using this model and basic combinatorics tools, we show that the expected size of a BN is $O\big(kn-k^2\ln n\big)$, the expected number of edges added when moralizing a BN is $O\big(k^2n-k^3\ln n\big)$, and the expected size of the Markov blanket is $O\big(k^2-k^3\ln n/n\big)$. We also show that the maximum size of a minimal d-separator is k; and that on the average, a revised version of the algorithm for finding a minimal d-separator by Acid and De Campos (1996) and Tian et al. (1998) has time complexity upper-bounded by $O\big(k^2(k^3+1)n-k^2\ln n\big)$. Non-Big-O estimates can be found in Section 4-6. In addition, we compare our estimates against BNs for analyzing gene expression data learned from the Gene Expression Omnibus database (Edgar et al., 2002; Friedman et al., 2000). The results suggest that this model provides a useful characterization of the BNs.

The BN distribution of our model has been used as a prior for BN learning, known as the order-modular prior (Friedman and Koller, 2000). To the best of our knowledge, the mathematical property of the model have yet received much attention. Most work on random graphs is about undirected graphs (Bollobás, 2001; Janson et al., 2000; Newman, 2009). In addition, a node in BNs usually has a limited number of parents (Ide and Cozman, 2002). The studies on random DAGs have not considered DAGs with this property (Barak and Erdös, 1984; Cohen et al., 2003; Dorogovtsev et al., 2001; Karrer and Newman, 2009; Pittel and Tungol, 2001).

Appearing in Proceedings of the 17^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2014, Reykjavik, Iceland. JMLR: W&CP volume 33.

2 THE MODEL

Inspired by the work of Barak and Erdös (1984) on random DAGs, we propose the following model for random BNs.

Definition 1. *(Random Bayesian Network)*
Given a set of nodes V, where $|V| = n$, and an integer k, let $G_{n,k}$ denote the random Bayesian network *for V where each node has at most k parents. Then let L denote a list of DAGs constructed as follows: Consider all possible orderings $(V, <)$ of the nodes in V. For each ordering $(V, <)$, every node v has zero to k parents v', where $v' < v$. Then $P(G_{n,k} = G) = n_G/|L|$, where n_G is the number of times that G appears in L.*

For example, consider $V = \{X, Y, Z\}$ and $k = 1$. In Definition 1, when $(V, <) = \langle X, Y, Z\rangle$ or $\langle X, Z, Y\rangle$, the DAGs constructed are

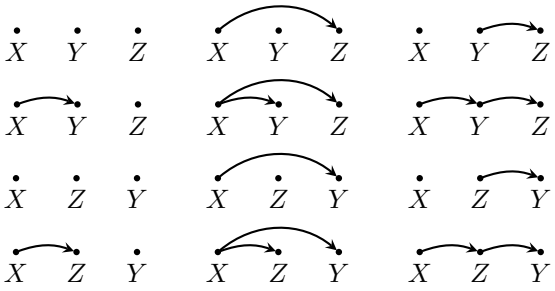


Figure 1: Given $V = \{X, Y, Z\}$ and $k = 1$, when $(V, <) = \langle X, Y, Z\rangle$ or $\langle X, Z, Y\rangle$, the DAGs constructed in Definition 1.

Note that there are six possible $(V, <)$. Let G be the edgeless DAG, that is, $X\ Y\ Z$. Since G appears once for each $(V, <)$, and each $(V, <)$ induces six possible DAGs, $P(G_{n,k} = G) = 1/6$. Then consider the DAG $G' : X \to Y \to Z$. Since G' only appears once when $(V, <) = \langle X, Y, Z\rangle$, $P(G_{n,k} = G') = 1/36$.

From the example above, it appears that $G_{n,k}$ is likely to contain isolated nodes. Section 7 shows that when n is large, $G_{n,k}$ contains no isolated nodes almost surely.

In addition, note that for a graph G in L, since there exists a $(V, <)$ such that the parents are always smaller than the children, G does not contain cycles, and is therefore a DAG. Moreover, since all possible $(V, <)$ are considered, L contains all the possible DAG configurations for $G_{n,k}$.

Now we discuss some fundamental properties of our random BN model. Given a $(V, <)$, consider distinct nodes v, v', pa, ch where $pa < v, v' < ch$. Recall that each node has at most k parents. Consequently if v is already a parent of ch, then v' is less likely to be a parent of ch. However since the constraint is only on the number of parents, whether pa is a parent of v is independent of whether pa is a parent of v'. We state the properties formally as the following theorems.

Theorem 1. *(Children Dependence)*
Given a random Bayesian network $G_{n,k}$ and an ordering $(V, <)$, for distinct nodes v, v', ch where $v, v' < ch$, the event that whether ch is a child of v and the event that whether ch is a child of v' are dependent.

Theorem 2. *(Parent Independence)*
Given a random Bayesian network $G_{n,k}$ and an ordering $(V, <)$, for distinct nodes v, v', pa where $pa < v, v'$, the event that whether pa is a parent of v and the event that whether pa is a parent of v' are independent.

3 MATHEMATICAL TOOLS

Now we discuss the combinatorics tools that we use to analyze our random BN model.

First we show the absorption and extraction identities of the binomial coefficients. Let i be a non-negative integer. Then

$$(i+1)\binom{n}{i+1} = n\binom{n-1}{i} \tag{1}$$

$$(n-i)\binom{n}{i} = n\binom{n-1}{i}. \tag{2}$$

Then we consider some fundamental tools in probabilistic combinatorics. Let $I_1, \ldots, I_i$ be random variables and I_n be a non-negative random variable that depends on n. Then

$$E[I_1 + \ldots + I_i] = E[I_1] + \ldots + E[I_i]. \tag{3}$$

$$\text{If } \lim_{n\to\infty} E[I_n] \to 0, \text{ then } I_n \text{ is zero almost surely.} \tag{4}$$

Equation (3) is known as the linearity of expectation. Note that $I_1, \ldots, I_i$ may be dependent. Consequently (3) is a valuable tool for tackling problems involving dependent events. Equation (4) can be intuitively understood as follows: Since random variable I_n is non-negative and its expected value approaches zero, it is almost always zero (Alon and Spencer, 2008).

In addition, we use integral to approximate summation (Graham et al., 1994). Let $f(r)$ be a smooth function defined for all reals r in $[m, n]$, then

$$\sum_{i=m}^{n} f(i) \approx \int_m^n f(r)\, dr.$$

In the following, when integral is used to approximate $\sum_{i=m}^{n} f(r)$ to a real number r', we write $\sum_{i=m}^{n} f(r) \sim r'$. In addition, approximations made without explanation are based on the fact that $k \ll n$.

Moreover, let I_A denote the indicator random variable of event A, that is, $I_A = 1$ when A occurs, and $I_A = 0$ when A does not. Note that $E[I_A] = P(A)$.

4 BASIC PROPERTIES

In this section, we study the probability that a given edge occurs, the expected BN size, and the expected number of parents for a node in our random BN model.

Recall that the definition of our model involves all possible $(V, <)$. Consider mapping each $(V, <)$ to integers between 1 and n. For example, for $(V, <) = \langle X, Y, Z\rangle$, map X, Y, Z to $1, 2, 3$. The mapping preserves the node ordering, and therefore can be used to represent $(V, <)$.

4.1 Edge Probability

In this subsection, we consider the probability that there exists edge $X \to Y$ for given X, Y in $G_{n,k}$.

4.1.1 Probability Estimation

Let A denote the event that there exists $X \to Y$, and let O denote the event that in $(V, <)$, X, Y map to integers x, y. Note that the edge exists only if $x < y$. Consequently

$$P(A) = \sum_{1 \le x < y \le n} P(A|O)P(O).$$

When $x < k$, Y does not have k nodes to serves as parents. Since $n \gg k$, we ignore these cases. Hence

$$P(A) \approx \sum_{k \le x < y \le n} P(A|O)P(O).$$

Clearly $P(O) = 1/\left(n(n-1)\right)$. Then note that $P(A|O)$ is the ratio of the number of DAGs where X is a parent of Y to the number of DAGs where X may or may not be a parent of Y when given $X < Y$. By Theorem 1 and 2, this ratio is the same as the ratio of possible parents of Y when X is a parent of Y to possible parents of Y when X may or may not be a parent of Y. When X is a parent, there are $y-2$ nodes, other than X, smaller than Y and may be parents of Y. Consequently there are $\sum_{i=0}^{k-1} \binom{y-2}{i}$ ways to choose parents of Y. When X may or may not be a parent, there are $\sum_{i=0}^{k} \binom{y-1}{i}$ ways to choose parents of Y. Hence

$$P(A|O) = \frac{\sum_{i=0}^{k-1} \binom{y-2}{i}}{\sum_{i=0}^{k} \binom{y-1}{i}}$$

A partial summation of the binomial coefficients such as the numerator and denominator in the equation above does not have a closed form (Graham et al., 1994). Therefore we now simplify this equation. Let $S_1 = \sum_{i=0}^{k-1} \binom{y-2}{i}$ and $S_2 = \sum_{i=0}^{k} \binom{y-1}{i}$. Then

$$\begin{aligned} S_1 &= \frac{1}{y-1} \sum_{i=0}^{k-1} (i+1) \binom{y-1}{i+1} \text{(by (1))} \\ &= \frac{1}{y-1} \sum_{i=0}^{k} i \binom{y}{i} \qquad (5) \\ S_1 &= \frac{1}{y-1} \sum_{i=0}^{k} (y-1-i) \binom{y-1}{i} - \binom{y-2}{k} \text{(by (2))} \\ &= S_2 - S_1 - \binom{y-2}{k} \text{(by (5))}. \end{aligned}$$

Consequently

$$P(A|O) = \frac{1}{2}\left(1 - \frac{\binom{y-2}{k}}{\sum_{i=0}^{k} \binom{y-1}{i}}\right).$$

Since when y is not small, $\binom{y-1}{k} + \binom{y-1}{k-1} = \binom{y}{k}$ are dominant in $\sum_{i=0}^{k} \binom{y-1}{i}$, we approximate $\sum_{i=0}^{k} \binom{y-1}{i}$ with $\binom{y}{k}$. Similarly, we approximate $y-1$ with y. As a result,

$$P(A|O) \approx \frac{k}{y} - \frac{k(k+1)}{2y^2}. \qquad (6)$$

Then by using integral to approximate summation,

$$P(A) \sim \frac{k}{n} - \frac{k(k+3)\ln n}{2n^2}.$$

4.1.2 Approximation Error

Now we give a naive analysis on the error of $P(A)$ estimation from approximating $\sum_{i=0}^{k} \binom{y-1}{i}$ with $\binom{y}{k}$, denoted as $e(A)$.

First consider when $y > 2k-1$. Note that

$$\sum_{i=0}^{k} \binom{y-1}{i} < \binom{y}{k} \frac{(y-k+1)(y-k+2)}{(y-k+1)(y-k+2) - k(k-1)}.$$

Please see the Appendix for details. Consequently the error in estimating $P(A|O)$, denoted as $e(A|O)$, is

$$\begin{aligned} e(A|O) &= \frac{1}{2}\left(\frac{\binom{y-2}{k}}{\binom{y}{k}} - \frac{\binom{y-2}{k}}{\sum_{i=0}^{k} \binom{y-1}{i}}\right) \\ &< \frac{k(k-1)(y-k)(y-k-1)}{2y(y-1)(y-k+1)(y-k+2)}. \end{aligned}$$

Since $n \gg k$,

$$\begin{aligned} e(A) &\approx \sum_{2k-1 \le x \le y \le n} e(A|O)P(O) \\ &\sim \frac{k(k-1)\ln n}{2n^2}. \end{aligned}$$

4.2 Expected Size and Number of Parents

Since $E[I_A]$ is the expected number of times $X \to Y$ exists for given X, Y, by (3), summing $E[I_A]$ over all possible X, Y gives the expected size of $G_{n,k}$:

$$\sum_{X,Y \in V} E[I_A] = \sum_{X,Y \in V} P(A) \approx kn - \frac{k(k+3)\ln n}{2}.$$

Then since each edge induces exactly one parent and there are n nodes, the expected number of parents a node has is approximately

$$k - \frac{k(k+3)\ln n}{2n}.$$

5 MORAL GRAPH AND MARKOV BLANKET ANALYSIS

Now we apply the analysis used in Section 4 to moral graphs and Markov blankets.

5.1 Moral Graph

In this subsection, we first consider the v-structure for given parents X, Y and child Z, that is, $X \to Z \leftarrow Y$ where there are no edges between X and Y. Then we consider the expected size increase of moralizing a BN.

Let B denote the event that there exists $X \to Z \leftarrow Y$ and there do not exist edges between X and Y in $G_{n,k}$. Let B_1 denote the event that there exists $X \to Z \leftarrow Y$, and let B_2 denote the event that there do not exist edges between X and Y. Let O denote the event that in $(V, <)$, X, Y, Z map to x, y, z. By Theorem 2, B_1 and B_2 are independent given O. Consequently

$$P(B) \approx \sum_{\substack{k \le x < y < z \le n \\ k \le y < x < z \le n}} P(B_1|O)P(B_2|O)P(O).$$

Consider when $k \le x < y < z \le n$. Note that edges between X and Y can only be $X \to Y$. Then similar to Subsection 4.1,

$$P(B_1|O) = \frac{\sum_{i=0}^{k-2} \binom{z-3}{i}}{\sum_{i=0}^{k} \binom{z-1}{i}} \approx \frac{k(k-1)}{4z^2}\left(3 - \frac{2(k+1)}{z}\right)$$

$$P(B_2|O) = 1 - \frac{\sum_{i=0}^{k-1} \binom{y-2}{i}}{\sum_{i=0}^{k} \binom{y-1}{i}} \approx 1 - \frac{k}{y} + \frac{k(k+1)}{2y^2}.$$

Please see the Appendix for details. When $k \le y < x < z \le n$, the case is essentially identical. As a result,

$$P(B) \sim \frac{3k(k-1)}{4n^2} - \frac{k(k-1)(7k+1)\ln n}{2n^3}.$$

Each v-structure induces exactly one edge to be added in moralization. Consequently the expected number of edges added when moralizing $G_{n,k}$ is

$$\sum_{\substack{\{X,Y\} \subset V \\ Z \in V}} E[I_B] \approx \frac{3k(k-1)n}{8} - \frac{k(k-1)(7k+1)\ln n}{4}.$$

5.2 Markov Blanket

The Markov blanket of a node consists of its parents, spouses, and children. Recall Subsection 4.2. Note that since each edge induces a parent and a child, the expected number of parents is identical to the expected number of children. In addition, since each v-structure induces two spouses, following Subsection 5.1, the expected number of spouses is approximately

$$\frac{3k(k-1)}{4} - \frac{k(k-1)(7k+1)\ln n}{2n}.$$

Then the expected size of the Markov blanket is approximately

$$\frac{k(3k+5)}{4} - \frac{k(7k^2-5k+2)\ln n}{2n}.$$

6 d-SEPARATOR ANALYSIS

In this section, we consider the problem of finding a minimal d-separator for given X, Y, that is, a d-separator such that none of its subsets d-separates X and Y. We study the maximum size of a minimal d-separator, and the time needed to find one.

6.1 Maximum Size

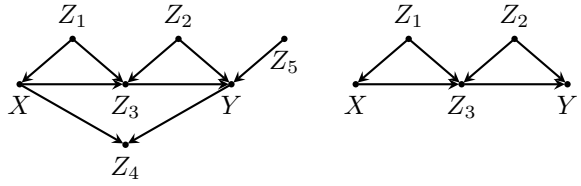
If there exists a d-separator for X and Y, at least one of the following two sets d-separates X and Y: the set consisting of the parents of X, and the set consisting of the parents of Y (Pearl, 1988). As a result, the maximum size of a minimal d-separator is k.

6.2 Complexity

We first review some results on this problem by Acid and De Campos (1996) and Tian et al. (1998). Let G be a DAG in $G_{n,k}$, and let G_A denote the subgraph induced by the ancestral set of $X \bigcup Y$. A minimal d-separator of X and Y only needs to d-separate them in G_A, and can be found by running two breadth-first searches on $(G_A)^m$, where m denotes the moral graph.

The results can be improved by exploiting the fact that a d-separator of X and Y only needs to d-separate them in G'_A, the subgraph of G_A consisting of paths between X and Y, as shown in Figure 2. Consequently

a minimal d-separator can be found by (1) construct G'_A, in $O(|G_A|)$ time (2) run two breadth-first searches on $(G'_A)^m$, in $O(|(G'_A)^m|)$ time. Then since G'_A is a subgraph of G_A, the time complexity for finding a minimal d-separator $O(|G_A|+|(G'_A)^m|) = O(|G_A|+|E_m|)$, where E_m is the edges added in G'_A moralization.



(a) DAG G. (b) DAG G'_A.

Figure 2: A DAG G and its G'_A.

Below, we show that the expected time complexity for finding a minimal d-separator in $G_{n,k}$, that is, $O(|G_A| + |E_m|)$, is asymptotically upper bounded by $k^3(k-1)^2n/20 + k^2n/4 - k^2 \ln n$.

6.2.1 Expected Size of G_A

Let Z, Z' be two nodes in $(V, <)$ where $Z < Z'$, and let S denote the event that edge $Z \to Z'$ is in G_A. Then

$$E[|G_A|] = \sum_{\{Z,Z'\}\subset V} E[I_S].$$

Let A denote the event that Z, Z' are ancestors of X in G, and let B denote that for Y. Let O denote the event that in $(V, <)$, X, Y, Z, Z' map to x, y, z, z'. Note that an event may not hold for some $(V, <)$, such as A when $x < z < z' < y$. Then

$$E[I_S] < \sum_{\substack{1\le y<z<z'<x\le n\\ 1\le z<y<z'<x\le n\\ 1\le z<z'<x<y\le n\\ 1\le z<z'<y<x\le n}} E[I_A|O]P(O) + \sum_{\substack{1\le x<z<z'<y\le n\\ 1\le z<x<z'<y\le n\\ 1\le z<z'<x<y\le n\\ 1\le z<z'<y<x\le n}} E[I_B|O]P(O).$$

First note that $\sum E[I_A|O]P(O) = \sum P(A|O)P(O)$. Then since event A states there exists path $Z \to Z' \to \ldots \to X$, consider such a path for a $(V, <)$, as below:

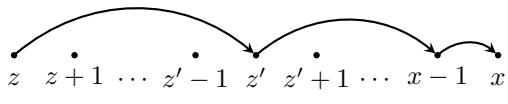


Figure 3: Path $Z \to Z' \to \ldots \to X$ for a $(V, <)$.

Note that the edges on the path appear independently. By (6), we approximate the probability that there exists $Z \to Z'$ with k/z', and that for $W_i \to X$ with k/x, where W_i are nodes between Z' and X. Then let p_i be the probability that there exist paths from Z' to W_i. Since the events that there exists $W_i \to X$ are dependent, and the events that there exist paths from Z' to W_i may also be dependent,

$$P(A|O) < \frac{k^2}{xz'}(p_1 + \ldots + p_{x-z'-1}) < \frac{k^2(x-z'-1)}{xz'}.$$

As a result,

$$\sum_{\substack{1\le y<z<z'<x\le n\\ 1\le z<y<z'<x\le n\\ 1\le z<z'<x<y\le n\\ 1\le z<z'<y<x\le n}} E[I_A|O]P(O) < \frac{k^2}{4n} - \frac{k^2 \ln n}{n^2}.$$

The second term is essentially the same. Consequently

$$E[|G_A|] < \frac{k^2 n}{4} - k^2 \ln n.$$

6.2.2 Expected Size of E_m

Let Z, Z' be two nodes in $(V, <)$ where $Z < Z'$, and let S' denote the event that edge $Z - Z'$ is added when moralizing G'_A. Then

$$E[|E_m|] = \sum_{\{Z,Z'\}\subset V} E[I_{S'}].$$

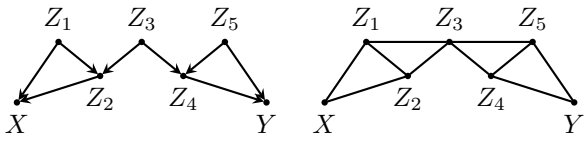
Edge $Z - Z'$ is added when there does not exist $Z \to Z'$, and there exist paths of the following types: (a) $X \to \ldots \to Z \to Z'' \leftarrow Z' \to \ldots \to Y$, where Z'' is an ancestor of Y (b) $X \leftarrow \ldots \leftarrow Z \to Z'' \leftarrow Z' \leftarrow \ldots \leftarrow Y$, where Z'' is an ancestor of X (c) $X \leftarrow \ldots \leftarrow Z \to Z'' \leftarrow Z' \to \ldots \to Y$, where Z'' is an ancestor of X or Y, or both.

To see this, first note that in (a) Z'' cannot be an ancestor of X or otherwise G contains a loop. Then consider path $X \leftarrow Z_1 \to Z_2 \leftarrow Z_3 \to Z_4 \leftarrow Z_5 \to Y$ in Figure 4(a). It appears to induce two edges $Z_1 - Z_3$ and $Z_3 - Z_5$ to be added, and yet is not any of the three types. Nevertheless, since in G'_A, nodes other than X and Y are ancestors of X or Y, there exist paths of the three types that induce $Z_1 - Z_3$ and $Z_3 - Z_5$, which in this case are $X \leftarrow Z_1 \to Z_2 \leftarrow Z_3 \to Z_4 \to Y$ and $X \leftarrow Z_2 \leftarrow Z_3 \to Z_4 \leftarrow Z_5 \to Y$.

By reasoning similar to that for $E[|G_A|]$,

$$E[|E_m|] < \frac{431k^3(k-1)^2n}{9216}.$$

Please see the Appendix for details.



(a) G'_A. (b) $(G'_A)^m$.

Figure 4: DAG G'_A and its $(G'_A)^m$.

7 ISOLATED NODES

In this section, we prove that $G_{n,k}$ contains no isolated nodes almost surely, that is, with probability 1.

Let A denote the event that a given node X is isolated, that is, in $(V, <)$, none of the nodes smaller than X is a parent of X, and none of the nodes greater than X is a child of X. Let A_1 denote the event that none of the nodes smaller than X is its parent, and let A_2 denote the event that none of the nodes greater than X is its child. Let O denote the event that in $(V, <)$, X maps to x. By Theorem 2, A_1 and A_2 are independent given O. Hence similar to Subsection 4.1,

$$P(A) \approx \sum_{x=k+1}^{n} P(A_1|O)P(A_2|O)P(O)$$
$$P(A_1|O) = \frac{1}{\sum_{i=0}^{k} \binom{x-1}{i}}.$$

Then by Theorem 2, whether a node greater than X is a child of X is independent of whether another such node is a child of X. By (6), we approximate the probability that there exists $X \to Y$ with k/y. Hence

$$P(A_2|O) \approx \prod_{i=x+1}^{n} 1 - \frac{k}{i}.$$

Then

$$P(A) < \frac{1}{n} \sum_{x=k+1}^{n} \frac{\prod_{i=x+1}^{n} 1 - \frac{k}{i}}{\binom{x}{k}} < \frac{1}{n} \sum_{x=k+1}^{n} \frac{k!}{(n-k+1)^k}$$
$$\sim \frac{k!}{(n-k+1)^k}.$$

As a result, the expected number of nodes that are isolated

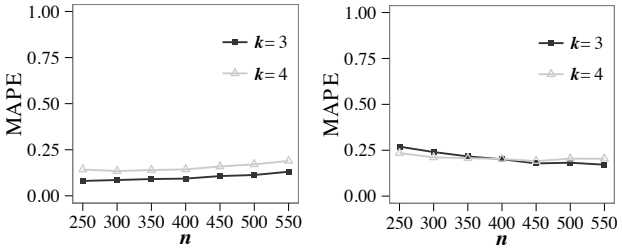
$$\sum_{X \in V} E[I_A] < \frac{k!n}{(n-k+1)^k}.$$

Since $k > 1$, this number approaches zero when $n \to \infty$. Hence by (4), $G_{n,k}$ contains no isolated nodes almost surely.

8 EXPERIMENTS

In this section, we compare our estimates of the expected size and the expected size increase of moralization against BNs learned from real data.

We use BN learning software Banjo (Hartemink, 2005) and the Gene Expression Omnibus database (Edgar et al., 2002) to learn BNs for analyzing gene expression data (Friedman et al., 2000). Banjo learns BN structures with the Bayesian Dirichlet scoring metric for a given maximum number of parents k. We learned about 30 BNs for $n = 250, 300, 350, 400, 450, 500, 550$ and $k = 3, 4$ respectively with an equivalent sample size that grows with k. Figure 5 shows the mean absolute percentage error (MAPE) of the estimates.



(a) The expected size. (b) The expected moralization size increase.

Figure 5: MAPE of the expected size and the expected size increase of moralization estimates.

The MAPE for the expected size is lower than 0.15 when $k = 3$, and lower than 0.2 when $k = 4$, though both slightly increase with n. The MAPE for the expected size increase for $n = 250$ is 0.27 when $k = 3$, and 0.23 when $k = 4$; and for $n = 550$ the MAPE drops to 0.17 when $k = 3$, and 0.20 when $k = 4$. The MAPEs suggest that our model provides a useful characterization of BNs with important applications. They also suggest that though our estimates are asymptotic ones, they may be applied to relatively small BNs.

9 CONCLUSION

We propose a simple model for large random BNs to study important properties of large BNs and BN algorithms. In this paper, we focus on the average case, that is, expectation, analysis. A natural next step is variance analysis. However, much of our expectation analysis relies on the linearity of expectation, and since variance is not linear, the analysis do not always apply to variance. Consequently we will explore variance analysis in future work.

Appendix

Approximation Error

Let $t = y - k$. When $y > 2k - 1$. Then

$$\begin{aligned}\frac{\sum_{i=0}^{k}\binom{y-1}{i}}{\binom{y}{k}} &= \frac{\binom{y}{k}+\binom{y}{k-2}+\binom{y}{k-4}\cdots}{\binom{y}{k}}\\ &< 1+\frac{k(k-1)}{(t+1)(t+2)}+\left(\frac{k(k-1)}{(t+1)(t+2)}\right)^2+\cdots\\ &\le \frac{(t+1)(t+2)}{(t+1)(t+2)-k(k-1)}.\end{aligned}$$

Moral Graph

Consider $P(B_1|O)$. Let $S_1 = \sum_{i=0}^{k-2}\binom{z-3}{i}$ and $S_2 = \sum_{i=0}^{k}\binom{z-1}{i}$. Let i be a positive integer. Note that

$$\binom{n}{i} = \binom{n-1}{i}+\binom{n-1}{i-1}. \tag{7}$$

Then similar to Subsection 4.1,

$$S_1 = \frac{1}{z-2}\sum_{i=0}^{k-1} i\binom{z-2}{i} \text{ (by (1))}, \tag{8}$$

$$\begin{aligned}S_1 &= \frac{1}{z-2}\sum_{i=0}^{k-2}(z-2-i)\binom{z-2}{i} \text{ (by (2))}\\ &= \frac{1}{z-1}\sum_{i=0}^{k-2}(z-1-i)\binom{z-1}{i}-S_1+\binom{z-3}{k-2}\text{(by (1, 2, 8))}\\ &= S_2-\binom{z}{k}-\frac{1}{z-1}\sum_{i=0}^{k-2} i\binom{z-1}{i}-S_1+\binom{z-3}{k-2}\text{(by (7))}.\end{aligned}$$

Let $S_3 = \sum_{i=0}^{k-2} i\binom{z-1}{i}$. Then

$$\begin{aligned}S_3 &= \sum_{i=0}^{k-3}(z-1-i)\binom{z-1}{i} \text{ (by (1) (2))}\\ &= -S_3+(z-1)\left(S_2-\binom{z}{k}-\binom{z-1}{k-2}+\binom{z-2}{k-3}\right)\text{(by (1, 7))}.\end{aligned}$$

Consequently

$$\frac{S_1}{S_2} = \frac{1}{4}\left(1+\frac{-\binom{z}{k}+\binom{z-1}{k-2}-\binom{z-2}{k-3}+2\binom{z-3}{k-2}}{\sum_{i=0}^{k}\binom{z-1}{i}}\right)$$

and again we approximate $\sum_{i=0}^{k}\binom{z-1}{i}$ by $\binom{z}{k}$. Then

$$P(B_1|O) \approx \frac{k(k-1)}{4z^2}\left(3-\frac{2(k+1)}{z}\right).$$

Expected Size of E_m

Let A denote the event that in G'_A there does not exist $Z \to Z'$ and there exists a path of Type (a). Similarly, let B, C denote that for Type (b), (c) paths. Then

$$\begin{aligned}E[I_{S'}] \le &\sum_{x<z<z'<z''<y} P(A|O)P(O)\\ &+\sum_{y<z<z'<z''<x} P(B|O)P(O)\\ &+\sum_{\substack{z<x<z'<z''<y\\ z<z'<x<z''<y\\ z<z'<z''<x<y\\ z<y<z'<z''<x\\ z<z'<y<z''<x\\ z<z'<z''<y<x}} P(C|O)P(O).\end{aligned}$$

Consider the first term. Recall Section 5. We approximate the probability that there exists $W \to Y \leftarrow W'$ with $3k(k-1)/(4y^2)$, and approximate the probability that v-structure $Z \to Z'' \leftarrow Z'$ exists with $3k(k-1)/(4z''^2)\,(1-k/z') = 3k(k-1)(z'-k)/(4z'z''^2)$. Then similar to Subsection 6.2.1,

$$P(A|O) < \frac{9k^3(k-1)^2(z'-k)(z-x-1)(y-z'-1)(y-z''-1)}{16y^2zz'z''^2}$$

and

$$\sum_{x<z<z'<z''<y} P(A|O)P(O) < \frac{5k^3(k-1)^2}{3072n^2}.$$

The case for the second term is the same as the first. Then consider the third term $\sum P(C|O)P(O)$. First note that the case for $\sum_{z<x<z'<z''<y} P(C|O)P(O)$ is essentially identical to $\sum_{x<z<z'<z''<y} P(A|O)P(O)$. Then when $z < z' < x < z'' < y$,

$$P(C|O) < \frac{9k^3(k-1)^2(z'-k)(x-z-1)(y-z'-1)(y-z''-1)}{16xy^2z'z''^2}$$

and

$$\sum_{z<z'<x<z''<y} P(C|O)P(O) < \frac{49k^3(k-1)^2}{18432n^2}.$$

Then consider when $z < z' < z'' < x < y$. Recall that Z'' is an ancestor of X or Y, or both. First consider when Z'' is an ancestor of X:

$$P(C|O) < \frac{9k^3(k-1)^2(z'-k)(x-z-1)(x-z''-1)(y-z'-1)}{16x^2yz'z''^2}.$$

Then consider when Z'' is an ancestor of Y:

$$P(C|O) < \frac{9k^3(k-1)^2(z'-k)(x-z-1)(y-z'-1)(y-z''-1)}{16xy^2z'z''^2}.$$

Consequently

$$\sum_{z<z'<z''<x<y} P(C|O)P(O) < \frac{251k^3(k-1)^2}{6144n^2}.$$

The remaining cases are the same as the above.

Acknowledgement

The authors would like to thank the reviewers, Professor Frederick Eberhardt, and Antti Hyttinen for their helpful comments. This research was supported in parts by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211 and #IIS-1018922, and ONR #N000-14-09-1-0665 and #N00014-10-1-0933.

References

S. Acid and L. M. De Campos. An algorithm for finding minimum d-separating sets in belief networks. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 3–10. Morgan Kaufmann Publishers Inc., 1996.

N. Alon and J. H. Spencer. *The probabilistic method*. Wiley, third edition, 2008.

A. B. Barak and P. Erdös. On the maximal number of strongly independent vertices in a random acyclic directed graph. *SIAM Journal on Algebraic Discrete Methods*, 5(4):508–514, 1984.

B. Bollobás. *Random graphs*. Cambridge University Press, second edition, 2001.

R. Cohen and S. Havlin. *Complex networks: structure, robustness and function*. Cambridge University Press, 2010.

R. Cohen, A. F. Rozenfeld, N. Schwartz, D. Ben-Avraham, and S. Havlin. Directed and non-directed scale-free networks. In *Statistical Mechanics of Complex Networks*, pages 23–45. Springer, 2003.

S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Giant strongly connected component of directed networks. *Physical Review E*, 64(2):025101, 2001.

R. Edgar, M. Domrachev, and A. E. Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002.

N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *Journal of computational biology*, 7(3-4):601–620, 2000.

Nir Friedman and Daphne Koller. Being bayesian about network structure. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 201–210, 2000.

R. L. Graham, D. E. Knuth, and Patashnik O. *Concrete mathematics: a foundation for computer science*. Addison-Wesley, second edition, 1994.

H. Guo and W. Hsu. A survey of algorithms for real-time bayesian network inference. In *AAAI/KDD/UAI02 Joint Workshop on Real-Time Decision Support and Diagnosis Systems*, 2002.

A. Hartemink. Banjo (bayesian network inference with java objects). 2005. URL `http://www.cs.duke.edu/~amink/software/banjo`.

J. S. Ide and F. G. Cozman. Random generation of bayesian networks. In *Advances in Artificial Intelligence*, pages 366–376. Springer, 2002.

S. Janson, T. Luczak, and A. Rucinski. *Random graphs*. Cambridge University Press, 2000.

B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, and M. West. Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, pages 388–400, 2005.

B. Karrer and M. E. J. Newman. Random graph models for directed acyclic networks. *Physical Review E*, 80(4):046110, 2009.

D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. The MIT Press, 2009.

M. E. J. Newman. *Networks: an introduction*. Oxford University Press, 2009.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

B. Pittel and R. Tungol. A phase transition phenomenon in a random directed acyclic graph. *Random Structures & Algorithms*, 18(2):164–184, 2001.

E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan. Computational solutions to large-scale data management and analysis. *Nature Reviews Genetics*, 11(9):647–657, 2010.

J. Tian, J. Pearl, and A. Paz. Finding minimal d-separators. Technical report, UCLA Cognitive Systems Laboratory, 1998.