Enumeration of Labelled Essential Graphs

Bertran Steinsky *
Technical University of Graz
Steyrergasse 30
8010 Graz
Austria
steinsky@finanz.math.tu-graz.ac.at

July 22, 2008

Abstract

We present two recursive enumeration formulas for the number of labelled essential graphs. The enumeration parameters of the first formula are the number of vertices, chain components, and cliques, while the enumeration parameters of the second formula are the number of vertices and cliques. Both formulas may be used to count the number of labelled essential graphs with given number of vertices.

1 Introduction

Directed acyclic graphs (DAGs or ADGs) are used to represent conditional independencies among random variables, e.g. in the field of Bayesian networks. There occurs the problem that different DAGs can represent the same conditional independence relations [5], i.e., they are Markov equivalent. This article deals with counting essential graphs, which can be identified with the equivalence classes of this equivalence relation.

For the definitions in this section we mainly follow Andersson[1], Harary[3], and Robinson[6]. A *digraph* G is a pair (V, E), where $V = V(G)$ is a finite set, and $E = E(G)$ is a subset of $(V \times V) \setminus \{(a, a) \mid a \in V\}$. Let $G \equiv (V, E)$ be a digraph. We write an arrow $a \rightarrow b \in G$ if $(a, b) \in E$ and $(b, a) \notin E$ and a line $a - b \in G$ if $(a, b) \in E$ and $(b, a) \in E$. If G contains no line then it is called *directed*. An *immorality of* G is a triple (a, b, c) where the induced subgraph $G_{\{a,b,c\}}$ is $a \rightarrow b \leftarrow c$. We define the *skeleton* G^u of a digraph $G \equiv (V, E)$

*The author is supported by the Austrian Science Fund (FWF), Project P-20412. Part of this research was supported by START-Project Y96-MAT of the Austrian Science Fund (FWF).

to be $G^u :\equiv (V, E^u)$, where $E^u = \{(a,b) \mid (a,b) \in E \text{ or } (b,a) \in E\}$. We note that E^u contains only lines. A *path* of length $n \geq 1$ from a to b in G is a sequence $(a_0, a_1, \ldots, a_n)$ such that $a_0, a_1, \ldots, a_n \in V$, $a_0 = a, a_n = b$, $a_i \neq a_j$ for $0 \leq i < j \leq n$, and $(a_{i-1}, a_i) \in E$ for $i = 1, \ldots, n$. A *cycle* in G is a sequence $(a_0, a_1, \ldots, a_n)$ such that $a_0, a_1, \ldots, a_n \in V$, $a_0 = a_n$, $a_i \neq a_j$ for $1 \leq i < j \leq n$, and $(a_{i-1}, a_i) \in E$ for $i = 1, \ldots, n$. A directed graph that contains no cycle is called a *directed acyclic graph (DAG)*. Two DAGs D_1 and D_2 are *graphically equivalent*, and we write $D_1 \sim D_2$, if they have the same skeleton and the same immoralities. We notice that $\sim$ is an equivalence relation and that it is known by Verma and Pearl[9] that two DAGs are Markov equivalent if and only if they are graphically equivalent. The *essential graph* D^* associated with a DAG D is the digraph

$$D^* = \cup(D' \mid D' \sim D),$$

i.e., the union over all DAGs that are graphically equivalent to D. A digraph is called an *essential graph* if it is the essential graph of some DAG.

G is called an *undirected graph* if $(u,v) \in E$ implies $(v,u) \in E$ for all $u, v \in V$. An undirected graph G is *connected* if for every pair of vertices a, b in G there is a path from a to b in G. A *connectivity component* of G is a maximal connected subgraph of G. We say a digraph G is *weakly connected* if its skeleton is connected and *strongly connected* if there is a path from every vertex to every other vertex. The *weak* or *strong components* of G are the maximal weakly or strongly connected subgraphs of G. G is a *chain graph* if its strong components are undirected connected graphs. The strong components of a chain graph G are called *chain components*. We say an undirected graph is *chordal* if every cycle of length $n \geq 4$ possesses a *chord*, i.e., two vertices connected by an edge, that is not part of the cycle. An arrow $a \to b \in G$ is *strongly protected* in G if $a \to b$ occurs in at least one of the four configurations in Figure 1 as an induced subgraph, where $c_1 \neq c_2$. We cite the following characterisation of essential graphs, which was

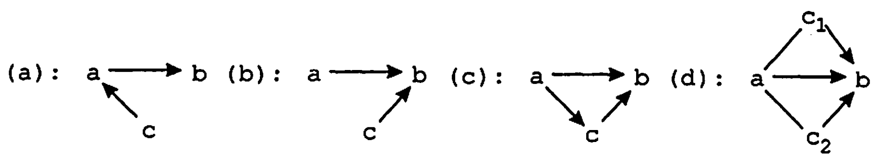

Figure 1: Strongly protected arrows.

given by Andersson *et* al.[1, Theorem 4.1], slightly reformulated.

Theorem 1 *(Andersson* et *al.[1, Theorem 4.1]) A digraph $G = (V, E)$ is equal to D^* for some DAG D, i.e., G is an essential graph, if and only if G satisfies the following four conditions:*

(i) G is a chain graph;

(ii) for every chain component τ of G, G_τ is chordal;

(iii) the configuration $a \to b - c$ does not occur as an induced subgraph of G;

(iv) every arrow $a \to b \in G$ is strongly protected in G.

We have already counted all labelled digraphs with the first three properties[8]. The following Lemma is the key to counting labelled digraphs with all four properties, i.e., essential graphs. Although it can be obtained from results of Gillispie[2], we prove it here, since it is the main tool for this article. A *clique* of a digraph G is an undirected subgraph S of G that is *complete*, i.e., there is an edge between all vertices $a, b \in S$.

Lemma 1 *Let G be an essential graph with n vertices and c cliques. Then the number of possibilities to add a set of arrows from vertices of G to a new vertex v_0 such that the resulting graph is essential is $2^n - c$, i.e., the cardinality of $\{E_0 \subseteq \{u \to v_0 \mid u \in V(G)\} \mid G_0 \equiv (V(G) \cup \{v_0\}, E(G) \cup E_0)$ is an essential graph$\}$ is $2^n - c$.*

Proof. The total number of possible connections from G to v_0 is 2^n. But we need to exclude connections where some of the arrows are not strongly protected. The *parents* of a vertex $v \in V$ are defined as the set $pa(v) = \{u \in V | u \to v\}$, where V is the vertex set of G. The parents of a vertex set $S \subseteq V$ is then $pa(S) = \bigcup_{v \in S} pa(v)$.

Let us assume that G is connected by arrows to the vertex v_0 in a way that some of the arrows are not strongly protected, i.e., the set $C = \{u \in V | u \to v_0 \text{ is not strongly protected}\} \subseteq pa(v_0)$ is not empty. It suffices to show that the induced subgraph of G on the vertex set C is a clique and that $pa(v_0) = C \cup pa(C)$, since then the connection is completely determined by the set C.

1. Let $u, v \in C$. If there is no edge between u and v, then the arrows from u and v to v_0 are strongly protected by configuration (b). If there is an arrow between u and v, then one of the arrows from u and v to v_0 is strongly protected by configuration (c). So there must be a line between u and v and therefore G induced on C is a clique.

2. First we show that $pa(C) \subseteq pa(v_0)$. Let $u \in pa(C)$ be a vertex that is not in $pa(v_0)$. There is a vertex $v \in C$ such that $u \to v$. By definition of v_0, there is no edge between u and v_0, which means that the arrow $v \to v_0$ is strongly protected by configuration (a), which is a contradiction. Therefore, we have $C \cup pa(C) \subseteq pa(v_0)$. To show that $pa(v_0) \subseteq C \cup pa(C)$ we prove that the set $N = pa(v_0) \setminus (C \cup pa(C))$ is empty. Let $u \in N$ and $v \in C$. If there is no edge between u and v, then $v \to v_0$ is strongly protected by configuration (b). If $v \to u$ then $v \to v_0$ is strongly protected by configuration (c). Since $u \to v$ is impossible by definition of N there must be a line $u - v$. This means that all vertices in C are connected to all vertices in N by a line. Condition (iii)

of Theorem 1 gives $pa(N) = pa(C)$. If N has only one element u, then $u \rightarrow v_0$ is not strongly protected by any configuration and this is a contradiction to the definition of C. If N has more than one element, then let $u, v \in N$. Any of the arrows $u \rightarrow v$ or $v \rightarrow u$ would produce a cycle in G, via a vertex in C. If there is no edge between u and v, then the arrow $w \rightarrow v_0$ for any vertex $w \in C$ would be strongly protected by configuration (d). Therefore the subgraph of G induced on the set $C \cup N$ is a clique and the arrow $u \rightarrow v_0$ is not strongly protected by any configuration, which is a contradiction as before. We finally have that N is empty. □

2 Enumeration of Labelled Chordal Graphs with Given Number of Vertices, Connectivity Components, and Cliques

In this section, our aim is to calculate $t(N, K, C)$, the number of labelled chordal graphs with N vertices, K connectivity components, and C cliques. For this purpose, we use methods and terminology of Wormald[10].

Let $G \equiv (V, E)$ be an undirected graph. A *j–clique* is a clique with j vertices. Let G be an undirected connected graph. A set of k vertices is called a *k-cutset* if G without this set is not connected. If there is no j-cutset for $j < k$ then we say G is *k-connected.* An exception is the complete graph with n vertices, which is defined to be k-connected if and only if $k \leq n$. The *connectivity* of G, denoted by $\kappa(G)$, is the maximal k for which G is k-connected.

Wormald[10, Lemma 1.1] showed that a graph G with connectivity k is chordal if and only if for every k-cutset W of G, W induces a clique of G and there are k-connected chordal graphs H_1 and H_2 with $H_1 \cap H_2 = W$ and $H_1 \cup H_2 = G$. We can also assume that $H_1 \neq G$ and $H_2 \neq G$.

Furthermore, Wormald associates with each connected chordal graph a vector called *maximal clique vector* of G. It is recursively defined as follows. If G is the complete graph with n vertices, then $mcv(G) = e_n$, where e_n is the infinite dimensional vector whose entries are all equal to 0 but the n-th is equal to 1. Otherwise, due to the previous characterisation of chordal graphs, $G = H_1 \cup H_2$, such that $H_1 \cap H_2$ is a $\kappa(G)$-cutset of G, where $H_1 \neq G \neq H_2$. Then

$$mcv(G) = mcv(H_1) + mcv(H_2) - e_{\kappa(G)}. \qquad (1)$$

For $\mathrm{i} = (i_1, i_2, \ldots)$ and $j \geq 1$ let

$$v_j(\mathrm{i}) = \sum_{k=1}^{\infty} \binom{k}{j} i_k.$$

A result of Wormald[10, Lemma 2.2] says that if $\mathbf{i} = mcv(G)$ then for $j \geq 1$, $v_j(\mathbf{i})$ is the number of j-cliques of G. Now we define

$$cl(\mathbf{i}) = \sum_{j=1}^{v_1(\mathbf{i})} v_j(\mathbf{i}),$$

which we call the *clique number* of G. We will sometimes write $cl(G)$. Let $a_{\mathbf{i}}$ be the number of labelled connected chordal graphs G for which $\mathbf{i} = mcv(G)$ and $MCV(N)$ be the set of all vectors that are mcv of some connected chordal graph with N vertices. Then

$$t(N, 1, C) = \sum_{\substack{\mathbf{i} \in MCV(N) \\ cl(\mathbf{i}) = C}} a_{\mathbf{i}},$$

i.e., the number of labelled connected chordal graphs with N vertices and C cliques. Using standard arguments, as they can be found for example at Harary[4, p. 7], we obtain

$$t(N, K, C) = \frac{1}{N} \sum_{r=1}^{N-1} r \binom{N}{r} \sum_{q=1}^{C} t(N-r, K-1, C-q) t(r, 1, q), \text{ for } K \geq 2.$$

To improve efficiency one should adjust the range of summation as follows.

$$t(N, K, C) = \frac{1}{N} \sum_{r=1}^{N-1} r \binom{N}{r} \sum_{q=\max(r, C-2^{N-r}+1)}^{\min(C-N+r, 2^r-1)} t(N-r, K-1, C-q) t(r, 1, q),$$

where $K \geq 2$.

3 Enumeration of Labelled Essential Graphs

3.1 Labelled Essential Graphs with Given Number of Vertices, Chain Components, and Cliques

Let $e(N, K, C)$ be the number of labelled essential graphs with N vertices, K chain components, and C cliques and $t(N, K, C)$ be the number of labelled chordal graphs with N vertices, K connectivity components, and C cliques where $N, K, C \geq 0$. We notice that $t(N, K, C)$ can be computed by the methods of Wormald[10].

Theorem 2 *$e(N, K, C)$ is equal to*

$$\sum_{n=1}^{N} \binom{N}{n} \sum_{k=1}^{K} (-1)^{k+1} \sum_{c=1}^{C} t(n, k, c) e(N-n, K-k, C-c) \left(2^{N-n} - C + c\right)^k, \quad (2)$$

where $e(0, 0, 0) = 1$.

Proof. We use some of Robinson's ideas([7] and [6, Proof of Theorem 1]) to derive the formula for $e(N, K, C)$. Let $1 \leq i \leq n$ and $E_i = E_i(N, K, C)$ be the set of all essential graphs with the vertex set $\{1, \ldots, N\}$, K chain components which are labelled from 1 to K, and C unlabelled cliques where the chain component labelled with i is a terminal component. Since every essential graph is a chain graph by condition (i) in Theorem 1, and every chain graph has at least one terminal chain component, we observe that $e(N, K, C)K! = |E_1 \cup \ldots \cup E_K|$ and by the inclusion-exclusion principle we have

$$|E_1 \cup \ldots \cup E_K| = \sum_{k=1}^{K} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq K} |E_{i_1} \cap \ldots \cap E_{i_k}|.$$

For fixed k we have $|E_1 \cap \ldots \cap E_k| = |E_{i_1} \cap \ldots \cap E_{i_k}|$ for any choice of $1 \leq i_1 < \ldots < i_k \leq K$. So we write

$$|E_1 \cup \ldots \cup E_K| = \sum_{k=1}^{K} (-1)^{k+1} \binom{K}{k} |E_1 \cap \ldots \cap E_k|.$$

Now, let G be a graph in $E_1 \cup \ldots \cup E_k$. We imagine putting those chain components of G that are labelled with i, where $1 \leq i \leq k$, on the right side and the other $K - k$ components on the left side. Let n be the number of vertices and c be the number of cliques of G on the right side, such that $1 \leq n \leq N$ and $1 \leq c \leq C$. The number of ways to choose the vertices on the right side is $\binom{N}{n}$. On the left side there is an essential graph with $N - n$ vertices, $C - c$ cliques, and $K - k$ chain components and on the right side there is a chordal graph with n vertices, c cliques, and k connectivity components. There are $e(N-n, K-k, C-c)(K-k)!$ possibilities for the essential graphs on the left side and $t(n, k, c)k!$ possibilities for the chordal graphs on the right side. (We notice that we consider the chain components on the left and the connectivity components on the right to be labelled.) Now, we observe that if there is an arrow $a \to b$ from a vertex a on the left side to a vertex b in a component γ on the right side then a has to point to all vertices of γ because of condition (iii) in Theorem 1. Now, let γ be any component on the right side. By Lemma 1 the number of possible connections from the left vertices to γ is $2^{N-n} - (C - c)$. This means that the number of possible sets of arrows from left to right is $\left(2^{N-n} - (C - c)\right)^k$. So we conclude that $|E_1 \cap \ldots \cap E_k|$ is equal to

$$k!(K-k)! \sum_{c=1}^{C} \sum_{n=1}^{N} \binom{N}{n} t(n, k, c) e(N-n, K-k, C-c) \left(2^{N-n} - C + c\right)^k,$$

finishing the proof. □

We notice that many of the summands in (2) are zero, but we provide an improved version (7) in Section 4.

3.2 Labelled Essential Graphs with Given Number of Vertices and Cliques

It is possible to get rid of the parameter K in the following way. Let $e(N, C)$ be the number of labelled essential graphs with N vertices and C cliques. That is,

$$
\begin{aligned}
& e(N,C) = \sum_{K=1}^{N} e(N,K,C) = \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{K=1}^{N} \sum_{k=1}^{K} (-1)^{k+1} \sum_{c=1}^{C} t(n,k,c) e(N-n, K-k, C-c) \left(2^{N-n} - C + c\right)^k \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{c=1}^{C} \sum_{k=1}^{N} \sum_{K=k}^{N} (-1)^{k+1} t(n,k,c) e(N-n, K-k, C-c) \left(2^{N-n} - C + c\right)^k \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{c=1}^{C} \sum_{k=1}^{N} \sum_{K=0}^{N-k} (-1)^{k+1} t(n,k,c) e(N-n, K, C-c) \left(2^{N-n} - C + c\right)^k \quad (3) \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{c=1}^{C} \sum_{k=1}^{N} \sum_{K=0}^{N-n} (-1)^{k+1} t(n,k,c) e(N-n, K, C-c) \left(2^{N-n} - C + c\right)^k \quad (4) \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{c=1}^{C} \sum_{k=1}^{N} (-1)^{k+1} t(n,k,c) \left(2^{N-n} - C + c\right)^k \sum_{K=0}^{N-n} e(N-n, K, C-c) \\
= & \sum_{n=1}^{N} \binom{N}{n} \sum_{k=1}^{N} (-1)^{k+1} \sum_{c=1}^{C} t(n,k,c) \left(2^{N-n} - C + c\right)^k e(N-n, C-c),
\end{aligned}
$$

where we assume $e(0, 0) = 1$. The equality between (3) and (4) holds because if $N - n < N - k$ and $N - n < K \leq N - k$ then $e(N - n, K, C - c) = 0$ and if $N - k < N - n$ then $k > n$ and therefore $t(n, k, c) = 0$.

3.3 Labelled Essential Graphs with Given Number of Vertices

Now, it is simple to calculate $e(N)$, the number of labelled essential graphs with N vertices, i.e.,

$$e(N) = \sum_{C=N}^{2^N-1} e(N, C).$$

See, for example, Harary[4, p. 8] to calculate the number of labelled connected essential graphs with N vertices from $e(N)$. Furthermore, as in Subsection 2, we can determine the number of labelled essential graphs with given number of vertices and connectivity components.

Table 1: The Number of Labelled Essential Graphs

N	$e(N)$
1	1
2	2
3	11
4	185
5	8782
6	1067825
7	312510571
8	212133402500
9	326266056291213
10	1118902054495975181
11	8455790399687227104576
12	139537050182278289405732939
13	4991058955493997577840793161279

4 Computation

In the calculation of $e(N, K, C)$ the summation indices should be chosen according to the following inequalities

$$k \leq n \leq c \leq 2^n - 1 \tag{5}$$

$$K - k \leq N - n \leq C - c \leq 2^{N-n} - 1 \tag{6}$$

since otherwise $t(n, k, c)e(N - n, K - k, C - c)$ is zero. Therefore, $e(N, K, C)$ equals

$$\sum_{n=1}^{N} \binom{N}{n} \sum_{k=k_1}^{k_2} (-1)^{k+1} \sum_{c=c_1}^{c_2} t(n, k, c)e(N - n, K - k, C - c) \left(2^{N-n} - C + c\right)^k, \tag{7}$$

where $k_1 = \max(1, K - N + n)$, $k_2 = \min(K - 1, n)$, $c_1 = \max(n, C - 2^{N-n} + 1)$, and $c_2 = \min(C - N + n, 2^n - 1)$. In the same way we formulate

$$e(N, C) = \sum_{n=1}^{N} \binom{N}{n} \sum_{k=1}^{n} (-1)^{k+1} \sum_{c=c_1}^{c_2} t(n, k, c) \left(2^{N-n} - C + c\right)^k e(N - n, C - c).$$

We note that in spite of these improvements the computation of $e(N)$ requires exponential time. Most of the time to calculate the numbers in Table 1 was taken by the computation of $t(n, k, c)$. As Wormald[10] noted, the number of arithmetic operations required for the calculation of the number of chordal graphs with n vertices, in an elementary implementation of his methods, is roughly 4^n times a constant.

5 Acknowledgement

We want to thank Steven B. Gillispie for helpful and inspiring correspondence and we are grateful to the referee for detailed comments, which improved this article.

References

[1] S. A. Andersson, D. Madigan, and M. D. Perlman, A Characterization of Markov Equivalence Classes for Acyclic Digraphs, Ann. Stat. 25, No 2 (1997) 502–541.

[2] S. B. Gillispie, Formulas for counting acyclic digraph Markov equivalence classes, Journal of Statistical Planning and Inference 136, No 4 (2006) 1410–1432

[3] F. Harary, R. Z. Norman, and D. Cartwright, Structural Models: An Introduction to the theory of graphs, Wiley, New York (1965).

[4] F. Harary, E. M. Palmer, Graphical Enumeration, Academic Press, New York (1973).

[5] D. Heckerman, D. Geiger, and D. M. Chickering,"Learning Bayesian networks: the combination of knowledge and statistical data", in: Uncertainty in Artificial Intelligence, Proceedings of the Tenth Conference

[6] R. W. Robinson, Counting labeled acyclic digraphs, in: F. Harary, ed., New Directions in the Theory of Graphs, Academic Press, New York (1973) 239–279.

[7] R. W. Robinson, Counting unlabeled acyclic digraphs, Lect. Notes Math. 622 (1977), 28–43.

[8] B. Steinsky, Enumeration of Labelled Chain Graphs and Labelled Essential Directed Acyclic Graphs, Discrete Mathematics 270 (2003), 267–278

[9] T. Verma, J. Pearl, "An algorithm for deciding if a set of observed independencies has a causal explanation", in: Uncertainty in Artificial Intelligence, Proceedings of the Eighth Conference (Morgan Kaufman, San Francisco, 1992), 323–330.