EFFICIENT CONSTRUCTION OF $\mathcal{B}$-ESSENTIAL GRAPHS

Luca La Rocca and Alberto Roverato

Dipartimento di Scienze Sociali, Cognitive e Quantitative
Università degli Studi di Modena e Reggio Emilia
Via Giglioli Valle 9, 42100 Reggio Emilia
(e-mail: {larocca.luca,roverato.alberto}@unimore.it)

ABSTRACT. *Essential graphs* and *largest chain graphs* are well-established graphical representations of equivalence classes of DAGs and chain graphs (CGs) respectively, especially useful in the context of structural learning. Recently, Roverato and La Rocca (2004) introduced the notion of a *labelled block ordering of vertices* $\mathcal{B}$ as a flexible tool for specifying subfamilies of CGs. In particular, both the family of DAGs and the family of "unconstrained" CGs are classes of $\mathcal{B}$*-consistent* CGs for the appropriate choice of $\mathcal{B}$. Equivalence classes of $\mathcal{B}$-consistent CGs are represented by means of $\mathcal{B}$*-essential graphs*. In this paper, we give an efficient procedure for the construction of $\mathcal{B}$-essential graphs that is based on the idea of a topological sorting of meta-arrows. In this way we also provide an efficient procedure for the construction of both largest chain graphs and essential graphs. The key feature of the proposed procedure is that every meta-arrow needs to be processed only once.

1 INTRODUCTION

A graphical Markov model for a random vector X_V is a family of probability distributions satisfying a collection of conditional independencies encoded by a graph with vertex set V. Every variable is associated with a vertex of the graph and the conditional independencies between variables are determined through a *Markov property*. Two different Markov properties are available, and in this paper we refer to the so called LWF Markov property; see Cowell *et. al.* (1999) for details.

Three main typologies of graphical models have been introduced: models for *directed acyclic graphs* (DAGs) models for *undirected graphs* (UGs) and models for *chain graphs* (CGs); the latter have both directed edges (arrows) and undirected edges (lines) but not semi-directed cycles. The introduction of CG models was motivated by substantive background knowledge on the existence of a *block ordering of variables* (Wermuth and Lauritzen (1990)); nevertheless, Lauritzen and Richardson (2002) critically discussed the role of block ordering of variables and, in particular, they pointed out that, in structural learning, restricting the attention to CG models with a particular pre-specified block ordering may preclude finding the most parsimonious model. A partial solution to this problem was provided by Roverato and La Rocca (2004) who introduced a more flexible way to specify subsets of CGs through a *labelled block ordering of vertices* $\mathcal{B}$ which identifies a subclass of $\mathcal{B}$*-consistent* CGs.

Several different CGs, with common vertex set, may be equivalent with respect to a given Markov property, in the sense that they define the same statistical model. Markov equivalence is an equivalence relation that induces a partition of the set of CGs into equivalence classes. From a statistical perspective, the point of interest is the statistical model. However, if a statistical model is represented by using an arbitrary graph in the respective *Markov equivalence* *class*, then the non-unique nature of the graphical description may result in difficulties. More specifically, structural learning procedures that deal with the space of CGs instead of the space of equivalence classes may face several problems concerning computational efficiency and the specification of prior distributions. On the other hand, the drawback of considering equivalence classes is that in this case most of the advantages deriving from the graphical representation of the models are lost, unless it is possible to characterise each equivalence class through a single representative CG. Frydenberg (1990) showed that the problem of choosing a representative for CGs has a natural solution because every equivalence class of CGs contains a *largest chain graph*, which is the graph with the largest amount of undirected edges within the equivalence class. The problem of Markov equivalence is also of interest for the class of all DAGs Markov equivalent to a given DAG $\mathcal{D}$, but in this case there is no DAG providing a natural representation of the class. Typically, equivalence classes of DAGs are characterised by means of the the smallest CG larger then every element of the class, that was called the *essential graph* by Andersson *et al.* (1997). Roverato and La Rocca (2004) dealt with the graphical characterisation of equivalence classes of $\mathcal{B}$-consistent CGs. They provided a representative, called the $\mathcal{B}$*-essential graph*, as well as a procedure for its construction.

The usefulness of a graphical representative is related to the availability of procedures to practically deal with it. In particular, a structural learning algorithm may require the repeated application of a procedure for the construction of the representative on the basis of any other graph in the Markov equivalence class. In this paper, we provide an efficient procedure for the construction of $\mathcal{B}$-essential graphs. Following the results of Roverato and La Rocca (2004) and Roverato (2005) the building-blocks of the procedure are *meta-arrows*, rather than single arrows, and we show that when meta-arrows are sorted according to a suitable topological ordering each of them needs to be considered only once. We also remark that, by properly choosing $\mathcal{B}$, this procedure can be used to construct both largest CGs and essential graphs.

The paper is organised as follows. In Section 2 we introduce the notion of $\mathcal{B}$-essential graph, together with some basic concepts and notation for graph and graphical model theory. In Section 3 we deal with $\mathcal{B}$-essential graph construction, first discussing the topological sorting of meta-arrows, then describing the proposed algorithm and showing its correctness.

2 $\mathcal{B}$-ESSENTIAL GRAPHS

In this section we review the theory of graphical models required in this paper. We introduce the notation we use as well as a few relevant concepts, but we omit the definitions of concepts such as *chain component* of a CG, (*minimal*) *complex* in a CG, *equivalence class*, *essential graph*, *largest CG*, *Markov equivalence*, *parent* of a vertex set and *semi-directed cycle* in a graph; we refer to Cowell *et al.* (1999) for a full account of the theory of graphs and graphical models, and to Roverato and La Rocca (2004) for the theory related to $\mathcal{B}$-consistent CGs.

We denote an arbitrary graph by $\mathcal{G} = (V, E)$, where V is a finite set of *vertices* and $E \subseteq V \times V$ is a set of *edges*. We say that $\alpha \in V$ and $\gamma \in V$ are joined by an *arrow* pointing at γ, and write $\alpha \rightarrow \gamma \in \mathcal{G}$, if $(\alpha, \gamma) \in E$ but $(\gamma, \alpha) \notin E$. We write $\alpha - \gamma \in \mathcal{G}$ if both $(\alpha, \gamma) \in E$ and $(\gamma, \alpha) \in E$, and say that there is an undirected edge, or *line*, between α and γ. If A and D are two chain components of a CG $\mathcal{G}$ such that there exist $\alpha \in A$ and $\delta \in D$ with $\alpha \rightarrow \delta \in \mathcal{G}$, we define the *meta-arrow* $A \rightrightarrows D$ as the set of all arrows in $\mathcal{G}$ pointing from A to D, i.e. we let $A \rightrightarrows D = \{\alpha \rightarrow \delta \in \mathcal{G} \mid \alpha \in A, \delta \in D\}$. A meta-arrow can be "merged" in the following way.

Definition 1. Let $\mathcal{G} = (V,E)$ be a CG and $A \rightrightarrows D$ one of its meta-arrows. The graph obtained from $\mathcal{G}$ by *merging* (the chain components connected by) $A \rightrightarrows D$ is the graph obtained from $\mathcal{G}$ by replacing every arrow $\alpha \rightarrow \delta \in A \rightrightarrows D$ with the corresponding line $\alpha — \delta$.

For a partition $V_1, \ldots, V_k$ of the vertex set V, a *labelled block ordering* $\mathcal{B}$ of V is defined as a sequence $\mathcal{B} = (V_1^{\ell_1}, \ldots, V_k^{\ell_k})$, shortly $\mathcal{B} = (V_i^{\ell_i})_{i=1}^k$, such that $\ell_i \in \{u,e,g\}$, $i = 1, \ldots, k$. A CG may, or may not, be consistent with a given labelled block ordering.

Definition 2. Let $\mathcal{B} = (V_i^{\ell_i})_{i=1}^k$ be a labelled block ordering of the vertex set V. We say that the CG $\mathcal{G} = (V,E)$ is $\mathcal{B}$*-consistent* when the following conditions hold:

(a) all edges joining vertices in different blocks of $\mathcal{B}$ are arrows pointing from blocks with lower numbering to blocks with higher numbering;
(b) for all i such that $\ell_i = u$, the subgraph $\mathcal{G}_{V_i}$ is an UG;
(c) for all i such that $\ell_i = d$, the subgraph $\mathcal{G}_{V_i}$ is a DAG.

Moreover, when $\mathcal{G}$ satisfies the above conditions with (c) relaxed to

(c′) for all i such that $\ell_i = d$, the subgraph $\mathcal{G}_{V_i}$ is a CG with decomposable chain components and there is no flag $\alpha \rightarrow \gamma — \delta \in \mathcal{G}$ such that $\gamma — \delta \in \mathcal{G}_{V_i}$

we say that $\mathcal{G}$ is *weakly* $\mathcal{B}$*-consistent*. We will call a $\mathcal{B}$-consistent CG a $\mathcal{B}$-CG and a weakly $\mathcal{B}$-consitent CG a $w\mathcal{B}$-CG for short.

A $\mathcal{B}$-consistent CG $\mathcal{G}$ identifies the class $[\mathcal{G}]^{\mathcal{B}}$ of all the $\mathcal{B}$-consistent CGs equivalent to $\mathcal{G}$. This can characterised by the $\mathcal{B}$*-essential graph* $\mathcal{G}^{\mathcal{B}}$, defined as the smallest $w\mathcal{B}$-CG larger than every element of $[\mathcal{G}]^{\mathcal{B}}$. Furthermore, $\mathcal{G}^{\mathcal{B}}$ is the unique $w\mathcal{B}$-CG equivalent to $\mathcal{G}$ with no $\mathcal{B}$-insubstantial meta-arrows (Roverato and La Rocca (2004)).

Definition 3. For a labelled block ordering $\mathcal{B} = (V_i^{\ell_i})_{i=1}^k$ of a vertex set V, let $\mathcal{G} = (V,E)$ be a $w\mathcal{B}$-CG and $A \rightrightarrows D$ a meta-arrow of $\mathcal{G}$. We say that the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-*insubstantial* in $\mathcal{G}$ when the following conditions hold:

(a) $A \cup D \subseteq V_i$ for some block V_i of $\mathcal{B}$;
(b) $\mathrm{pa}_{\mathcal{G}}(D) \cap A$ is complete;
(c) $\mathrm{pa}_{\mathcal{G}}(D) \setminus A \subseteq \mathrm{pa}_{\mathcal{G}}(\alpha)$, for all $\alpha \in \mathrm{pa}_{\mathcal{G}}(D) \cap A$;
(d) $\ell_i = d$ and $\mathrm{pa}_{\mathcal{G}}(D) \setminus A = \mathrm{pa}_{\mathcal{G}}(\alpha)$, for all $\alpha \in \mathrm{pa}_{\mathcal{G}}(D) \cap A$.

Roverato and La Rocca (2004) showed that the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$ can be constructed by successively merging $\mathcal{B}$-insubstantial meta-arrows of $\mathcal{G}$, thus obtaining a sequence of equivalent $w\mathcal{B}$–CGs, until no $\mathcal{B}$-insubstantial meta-arrow is left. However, this procedure is not efficient because, when a meta-arrow is merged, previously $\mathcal{B}$-substantial meta-arrows may become $\mathcal{B}$-insubstantial. Consequently, a meta-arrow that is found to be $\mathcal{B}$-substantial needs to be checked again in the following (possibly several times) in order to verify whether its status has changed or not. To cope with this inefficiency, we provide in the next section a procedure in which every meta-arrow needs to be checked only once.

3 $\mathcal{B}$-ESSENTIAL GRAPH CONSTRUCTION

In this section, we first introduce a *topological ordering* for the meta-arrows of a $\mathcal{B}$-CG $\mathcal{G}$, and discuss the issue of sorting the meta-arrows of $\mathcal{G}$ according to this ordering, then we illustrate an algorithm for the construction of $\mathcal{G}^{\mathcal{B}}$ that takes advantage of considering the meta-arrows of $\mathcal{G}$ in the suggested topological order.

3.1 TOPOLOGICAL SORTING OF META-ARROWS

Let A and D be two distinct chain components of a CG $\mathcal{G}$. We will say that *A precedes D*, and write $A \prec D$, when there exists a sequence $A = C_0, C_1, \ldots, C_r = D$, $r \geq 1$, of chain components of $\mathcal{G}$ such that $C_{i-1} \rightrightarrows C_i \in \mathcal{G}$, for all $i = 1, \ldots, r$. It is immediate to check that $\prec$ is a *partial ordering* of the chain components of $\mathcal{G}$ and we will call it their *natural topological ordering*. Given a CG $\mathcal{G}$, its chain components can be sorted according to their natural topological ordering (i.e. given a well-ordering that extends the partial ordering $\prec$) by applying the well known *topological sort* algorithm (see for example Cowell *et. al.* (1999)) to the DAG of chain components of $\mathcal{G}$. Note that, in general, the well-ordering extending $\prec$ is not unique.

The natural topological ordering of the chain components of a CG $\mathcal{G}$ induces two possible *lexicographical* orderings on the meta-arrow of $\mathcal{G}$, depending on whether the heads or the tails of the meta-arrows are compared first. In particular, to our aims, we find useful to compare heads first, that is to say that $A \rightrightarrows D$ *precedes* $C \rightrightarrows F$, and write $A \rightrightarrows D \prec C \rightrightarrows F$, when $D \prec F$ or $D = F$ and $A \prec C$. We will refer to this partial ordering of the meta-arrows of $\mathcal{G}$ as to their *heads-first topological ordering*, and we remark that it satisfies the two following conditions:

(i) if $A \rightrightarrows C \in \mathcal{G}$ and $C \rightrightarrows D \in \mathcal{G}$, then $A \rightrightarrows C \prec C \rightrightarrows D$;
(ii) if $A \rightrightarrows D \in \mathcal{G}$, $A \rightrightarrows C \in \mathcal{G}$ and $C \rightrightarrows D \in \mathcal{G}$, then $A \rightrightarrows C \prec A \rightrightarrows D \prec C \rightrightarrows D$.

A well-ordering of the meta-arrows of $\mathcal{G}$ that extends their heads-first topological ordering can be obtained by visiting the edges of the DAG of chain components of $\mathcal{G}$, after its vertices (i.e. the chain components of $\mathcal{G}$) have been, in turn, sorted. A similar sorting algorithm was given by Chickering (1995) for the edges of a DAG $\mathcal{D}$, as a first step in identifying the arrows belonging to the essential graph of $\mathcal{D}$. In the following, the topological sorting of the meta-arrows of a CG $\mathcal{G}$ according to $\prec$ will be performed by an algorithm `sortMetaArrows` which, due to space reasons, we do not give explicitly.

3.2 ALGORITHM FOR THE CONSTRUCTION OF $\mathcal{G}^{\mathcal{B}}$

Let $\mathcal{G} = (V, E)$ be an arbitrary $\mathcal{B}$-CG. Algorithm 1 takes $\mathcal{G}$ as input and gives as output the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$. It is evident that every meta-arrow of $\mathcal{G}$ is considered only once. What has to be shown is that, given that the meta-arrows of $\mathcal{G}$ are first sorted by means of the algorithm `sortMetaArrows`, this is enough to find $\mathcal{G}^{\mathcal{B}}$. In other words, we need to show that Algorithm 1 is correct. This is guaranteed by Theorem 1, whose proof takes advantage of the following lemma.

Lemma 1. *Let $A \rightrightarrows D$ be the meta-arrow considered on line 5 of Algorithm 1; then $D = T_i$.*

Remark. In the light of Lemma 1, line 5 of Algorithm 1 simplifies to

let $A \rightrightarrows T_i$ be the unique meta-arrow of $\mathcal{H}$ such that $R_i \subseteq A$

Algorithm 1 Pseudo-code for $\mathcal{B}$-essential graph construction
Input: a $\mathcal{B}$-CG $\mathcal{G} = (V, E)$
Output: the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$

let $(R_i \rightrightarrows T_i, i = 1, \ldots, m) = \text{sortMetaArrows}(\mathcal{G})$
let $\mathcal{H} = \mathcal{G}$
for $i = 1$ to m **do**
  if the arrows of $R_i \rightrightarrows T_i$ are not undirected edges in $\mathcal{H}$ **then**
    let $A \rightrightarrows D$ be the unique meta-arrow of $\mathcal{H}$ such that $R_i \subseteq A$ and $T_i \subseteq D$
    if the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-insubstantial in $\mathcal{H}$ **then**
      merge $A \rightrightarrows D$ in $\mathcal{H}$
    end if
  end if
end for
return $\mathcal{H}$

Proof. Assume by contradiction that $A \rightrightarrows D$ is considered with $T_i \subset D$. Then, there exists another chain component S of $\mathcal{G}$ such that $S \subset D$ and either $S \rightrightarrows T_i \in \mathcal{G}$ or $T_i \rightrightarrows S \in \mathcal{G}$. As the meta-arrow of $\mathcal{G}$ joining S and T_i was merged at a previous iteration of the algorithm, the possibility $T_i \rightrightarrows S \in \mathcal{G}$ is ruled out by ordering condition (i). Therfore, it holds that $S \rightrightarrows T_i \in \mathcal{G}$. Moreover, the chain components R_i and S are adjacent in $\mathcal{G}$, otherwise any given arrow of $S \rightrightarrows T_i$ would form a minimal complex in $\mathcal{G}$ with each arrow of $R_i \rightrightarrows T_i$ and this would make impossible for S and T_i to be included in the same chain component D of the equivalent $w\mathcal{B}$-CG $\mathcal{H}$. As the possibility $S \rightrightarrows R_i \in \mathcal{G}$ would give a semi-directed cycle in $\mathcal{H}$, it necessarily holds that $R_i \rightrightarrows S \in \mathcal{G}$. However, this results in a contradiction, as then $R_i \rightrightarrows T_i$ precedes $S \rightrightarrows T_i$ by ordering condition (ii).

Theorem 1. *For all input $\mathcal{B}$-CG $\mathcal{G}$, the output of Algorithm 1 is the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$.*

Proof. It is enough to show, thanks to the results by Roverato and La Rocca (2004), that whenever $A \rightrightarrows T_i$ is not merged in $\mathcal{H}$ then all of its arrows belong to $\mathcal{G}^{\mathcal{B}}$: we do this by finite induction on the iteration counter i.

Let $R_i \rightrightarrows T_i$ be the first meta-arrow of $\mathcal{G}$ considered by Algorithm 1 such that the arrowhead of $A \rightrightarrows T_i$ is not $\mathcal{B}$-insubstantial in $\mathcal{H}$. According to Definition 3, this can be for one or more of the following reasons:

(1) $A \subseteq V_j$ and $T_i \subseteq V_h$ with $j < h$;
(2) $\text{pa}_{\mathcal{H}}(T_i) \cap A$ is not complete;
(3) $\exists \alpha \in \text{pa}_{\mathcal{H}}(T_i) \cap A$ such that $\exists \gamma \in \text{pa}_{\mathcal{H}}(T_i) \setminus A$ not belonging to $\text{pa}_{\mathcal{H}}(\alpha)$;
(4) $\exists \alpha \in \text{pa}_{\mathcal{H}}(T_i) \cap A$ such that $\exists \gamma \in \text{pa}_{\mathcal{H}}(\alpha)$ not belonging to $\text{pa}_{\mathcal{H}}(T_i) \setminus A$ and $A \cup T_i$ is part of a block labelled with 'd'.

If condition (1) holds, then all arrows of $A \rightrightarrows T_i$ belong to $\mathcal{G}^{\mathcal{B}}$ because of the $\mathcal{B}$-consistency of $\mathcal{G}^{\mathcal{B}}$. If condition (2) holds, then there exist α and α' belonging to A such that (α, T_i, α') is a complex. Therefore, there are arrows of $A \rightrightarrows T_i$ belonging to a minimal complex and all arrows of $A \rightrightarrows T_i$ necessarily belong to $\mathcal{G}^{\mathcal{B}}$. If condition (3) holds, then either $\alpha \to \gamma \in \mathcal{H}$ or α and γ are not adjacent. The first case is impossible, as by ordering condition (ii) the meta-arrow of $\mathcal{G}$ including $\alpha \to \gamma$ would have been considered before $R_i \rightrightarrows T_i$ and thus merged.

Note the role played here by Lemma 1. So α and γ are not adjacent and (α, T_i, γ) is a complex. Therefore, there is an arrow of $A \rightrightarrows T_i$ which is part of a minimal complex and all arrows of $A \rightrightarrows T_i$ necessarily belong to $\mathcal{G}^{\mathcal{B}}$. Finally, if condition (4) holds, let $C_\alpha \subseteq A$ be the chain component of $\mathcal{G}$ to which α belongs. Then, as $C_\alpha \prec T_i$, the meta-arrow of $\mathcal{G}$ including $\gamma \rightarrow \alpha$ was considered before $R_i \rightrightarrows T_i$ and thus merged. Therefore, condition (4) cannot hold.

Now let $R_i \rightrightarrows T_i$ be any meta-arrow of $\mathcal{G}$ considered by Algorithm 1 such that the arrowhead of $A \rightrightarrows T_i$ is not $\mathcal{B}$-insubstantial in $\mathcal{H}$, and assume that the thesis holds for all meta-arrows of $\mathcal{G}$ considered before. Clearly, conditions (1) and (2) can be dealt with as above. The same happens for condition (3), except for the case $\alpha \rightarrow \gamma \in \mathcal{H}$. In this case, however, ordering condition (ii) implies (note the role of Lemma 1) that the meta-arrow of $\mathcal{G}$ including $\alpha \rightarrow \gamma$ was considered before (and not merged). Therefore, the inductive hypothesis gives $\alpha \rightarrow \gamma \in \mathcal{G}^{\mathcal{B}}$ and all arrows of $A \rightrightarrows T_i$ belong to $\mathcal{G}^{\mathcal{B}}$ otherwise the latter would possess a semi-directed cycle. Finally, if condition (4) holds, let τ be any son of α in T_i. It holds that γ and τ are not adjacent, because $\tau \rightarrow \gamma \in \mathcal{H}$ would imply a semi-directed cycle in $\mathcal{H}$. Hence $\alpha \rightarrow \tau \in \mathcal{G}^{\mathcal{B}}$ because the alternative $\alpha - \tau \in \mathcal{G}^{\mathcal{B}}$, given that $\gamma \rightarrow \alpha \in \mathcal{G}^{\mathcal{B}}$ by the inductive hypothesis, would imply the presence of the flag $\gamma \rightarrow \alpha - \tau$ in the $w\mathcal{B}$-CG $\mathcal{G}^{\mathcal{B}}$, which is impossible.

ACKNOWLEDGEMENTS

Partial financial support was given by MIUR, Rome (PRIN 03).

REFERENCES

ANDERSSON, S.A., MADIGAN, D., PERLMAN, M.D. (1997): A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics, 25*, 505–541.

CHICKERING, D.M. (1995): A transformational characterization of equivalent Bayesian network structures. In: P. Besnard and S. Hanks: (Eds.) *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, San Francisco, 87–98.

COWELL, R.G., DAWID, A.P., LAURITZEN, S.L., SPIEGELHALTER, D.J. (1999): *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York.

FRYDENBERG M. (1990): The chain graph Markov property. *Scandinavian Journal of Statistics*, 17, 333–353.

LAURITZEN, S.L., RICHARDSON, T.S. (2002): Chain graph models and their causal interpretation (with discussion). *Journal of the Royal Statistical Society, Series B, 64*, 321–361.

ROVERATO, A. (2005): A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs. *Scandinavian Journal of Statistics*, 32, 295–312.

ROVERATO, A., LA ROCCA, L. (2004): On block ordering of variables in graphical modelling. *Research Report n. 29-04 of the Department of Social, Cognitive and Quantitative Sciences*, Reggio Emilia, October 2004 (submitted).

WERMUTH, N., LAURITZEN, S.L. (1990): On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *Journal of the Royal Statistical Society, Series B, 52*, 21–72.