PROBABILITY DISTRIBUTIONS WITH SUMMARY GRAPH STRUCTURE

By Nanny Wermuth*

Mathematical Statistics at Chalmers/ University of Gothenburg, Sweden

A joint density of many variables may satisfy a possibly large set of independence statements, called its independence structure. Often the structure of interest is representable by a graph that consists of nodes representing variables and of edges that couple node pairs. We consider joint densities of this type, generated by a stepwise process in which all variables and dependences of interest are included. Otherwise, there are no constraints on the type of variables or on the form of the generating conditional densities. For the joint density that then results after marginalising and conditioning, we derive what we name the summary graph. It is seen to capture precisely the independence structure implied by the generating process, it identifies dependences which remain undistorted due to direct or indirect confounding and it alerts to such, possibly severe distortions in other parametrizations. Summary graphs preserve their form after marginalising and conditioning and they include multivariate regression chain graphs as special cases. We use operators for matrix representations of graphs to derive matrix results and translate these into special types of path.

1. Introduction.

Graphical Markov models are probability distributions defined for a $d_V \times 1$ random vector variable Y_V whose component variables may be discrete or continuous and whose joint density f_V satisfies independence statements captured by an associated graph. One such type of graph has been introduced as a multivariate regression chain graph by Cox and Wermuth (1993, 1996), see also Section 3.1 below.

With such graphs, each independence constraint is specified for an ordered sequence of single or joint response variables so that – in the case of a joint Gaussian distribution – it implies a zero parameter in a univariate or multivariate linear regression model. For discrete random variables all multivariate regression graph models are smooth, see Drton (2008), i.e. they are curved exponential families, see e.g. Cox (2007), Section 6.8.

*Supported in part by the Swedish Research Society via the Gothenburg Stochastic Center and by the Swedish Strategic Fund via the Gothenburg Math. Modelling Center

AMS 2000 subject classifications: Primary 62H99; secondary 62H05, 05C50

Keywords and phrases: Concentration graph, Directed acyclic graph, Endogenous variables, Graphical Markov model, Independence graph, Multivariate regression chain graph, Partial closure, Partial inversion, Triangular system.

This type of chain graph is well suited for modeling developmental processes, such as in panel studies which provide longitudinal data on a group of individuals, termed the 'panel', about whom information is collected repeatedly, say over years or decades, and in studies of direct and indirect effects of possible causes on joint responses, see Cox and Wermuth (2004).

A directed acyclic graph is an important special case of a chain graph. It arises in a stepwise generating process whenever each response variable is univariate, see also Section 2.1 below. In this paper, we consider both directed acyclic graphs and multivariate regression chain graphs as representing independence structures of corresponding data generating processes. Summary graphs are introduced to detect consequences of such a data generating process after having marginalized and conditioned on some of the variables. They include multivariate regression chain graphs in a subclass.

As we shall show, each summary graph may be derived from a directed acyclic graph in node set V by marginalising over a subset M of V and conditioning on a disjoint subset C of V. The new node set is $N = V \setminus \{M, C\} = (u, v)$ where each node in v and no node in u is in the past of the conditioning set C. A corresponding factorization of the derived density is, written in a condensed notation of node sets,

$$f_{N|C} = f_{u|vC} f_{v|C} \,. \tag{1.1}$$

The summary graph corresponding to (1.1) has three types of edge, only undirected, full edges, $i \frac{\quad}{} k$, within v, only arrows $i \prec\!\!-\!\!- k$ starting in v and pointing to u, but within u possibly undirected, dashed lines, $i - - - k$ and arrows, $i \prec\!\!-\!\!- k$ so that one type of double edge can arise within u: $i \overset{\prec\!-\!-}{- - -} k$.

Summary graphs capture each independence implied by the generating graph for $f_{N|C}$ and no other independences. They preserve their form after marginalising and conditioning. These two important properties are shared by two other types of graph, the MC-graphs of Koster (2002) and the ancestral graphs of Richardson and Spirtes (2002). In fact, the independence structures captured by an MC-graph, a summary graph and a maximal ancestral graph coincide whenever the same sets of marginalizing nodes M and of conditioning nodes C are given for a generating directed acyclic graph. This follows from their so-called global Markov properties.

In special cases, the three types of graph may be identical, but in general, the MC-graph cannot be recovered from the summary graph nor the summary graph from the maximal ancestral graph. By contrast, algorithms are available to obtain an independence equivalent summary graph from a MC-graph, Sadeghi (2008), and an independence equivalent maximal ancestral graph from a summary graph, see Section 3.2 below.

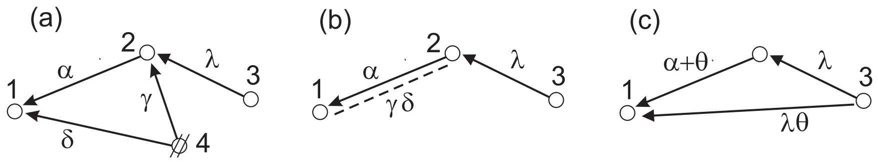
MC-graphs contain up to four types of edge coupling a given variable pair so that even for Gaussian distributions, they may correspond frequently to under-identified models, i.e. contain parameters that cannot be derived from the set of sufficient statistics. By contrast, every Gaussian maximal ancestral graph model contains three types of edge, never a double edge and deleting an arrow is equivalent to introducing an additional independence constraint so that parameters associated with $i \prec\!\!-\!\!- k$ measure a conditional dependence of Y_i on Y_k in $f_{u|vC}$.

This dependence may often differ qualitatively from the generating dependence of Y_i on Y_k in f_V, i.e. it may change the sign and remain a strong dependence. If this remains undetected, one would come to qualitatively wrong conclusions when interpreting the parameters measuring conditional dependence of Y_i on Y_k in $f_{u|vC}$. A summary graph provides the tool for detecting whether and for which of the generating dependences $i \prec\!\!-\!\!- k$, such distortions can occur due to direct or indirect confounding, see Wermuth and Cox (2008). We illustrate this here with two small examples where marginalising is represented by a crossed out node, $\not\!\circ$.

For a joint Gaussian distribution, the distortions are compactly described in terms of regression coefficients for variables Y_i standardized to have mean zero and variance one. For Figure 1a), the generating equations be

$$(1.2)\quad Y_1 = \alpha Y_2 + \delta\, Y_4 + \varepsilon_1, \quad Y_2 = \lambda Y_3 + \gamma Y_4 + \varepsilon_2, \quad Y_3 = \varepsilon_3, \quad Y_4 = \varepsilon_4,$$

FIG 1. *a) Generating graph for Gaussion relations in standardized variables, leading for variable Y_4 unobserved to b) the summary graph and c) the maximal ancestral graph for the observed variables; with the generating dependences as attached to re the arrows in a), implied are as simple correlations $r_{12} = \alpha + \gamma\delta, r_{13} = \alpha\lambda, r_{23} = \lambda$ and $\theta = \gamma\delta/(1-\lambda^2)$.*



With residuals ε_i assumed to have zero means and to be uncorrelated, the equations of the summary graph model that result from (1.2) for Y_4 unobserved, have one pair of correlated residuals

$$Y_1 = \alpha Y_2 + \eta_1, \quad Y_2 = \lambda Y_3 + \eta_2, \quad Y_3 = \eta_3,$$

$$\eta_1 = \delta\, Y_4 + \varepsilon_1, \quad \eta_2 = \gamma Y_4 + \varepsilon_2, \quad \eta_3 = \varepsilon_3, \quad \mathrm{cov}(\eta_1, \eta_2) = \gamma\delta.$$

The equation parameters to the Gaussian maximal ancestral graph associated with Figure 1c) are instead defined via

$$E(Y_1|Y_2 = y_2, Y_3 = y_3), \quad E(Y_2|Y_3 = y_3),$$

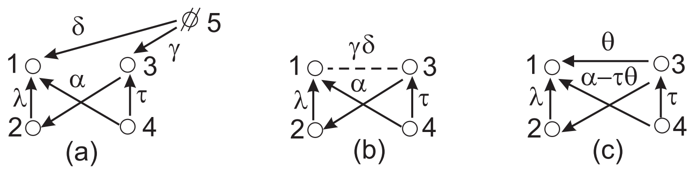
in this case, with all residuals in the recursive equations being uncorrelated. The generating dependence α is retained in the summary graph model.

The parameter for the dependence of Y_1 on Y_2 in the maximal ancestral graph model, expressed in terms of the generating parameters of Figure 1a), is $\alpha + \gamma\delta/(1-\lambda^2)$. The summary graph is here a graphical representation of the simplest type of an instrumental variable model, used in econometrics, see Sargan (1958), to separate a direct confounding effect, here $\gamma\delta$, from the dependence of interest, here α.

In general, the possible distortion, due to direct confounding occurring in parameters for dependence in maximal ancestral graph models, is recognized in the corresponding summary graph by a double edge $i \overset{\prec}{-\,-\,-} k$, see Wermuth and Cox (2008).

In the following second example for Gaussian standardized variables, there is no direct confounding of the generating dependence α but there is indirect confounding of α. The coefficient attached to $2 \prec\!\!-\!\!- 3$ is the simple correlation r_{23}. To simplify the figures, r_{23} is not displayed in the three graphs of Figure 2. The generating graph in Figure 2a) is directed and acyclic so that the corresponding linear equations, defined implicitly by Figures 2a) have uncorrelated residuals. The summary graph in Figure 2b) shows with a dashed line an induced association for pair $(1, 3)$, due to marginalising over Y_5.

FIG 2. *a) Generating graph for linear relations in standardized variables, leading for variable Y_5 unobserved to b) the summary graph and c) the maximal ancestral graph for the observed variables; with the generating dependences as attached to the arrows in the a), implied are $\theta = \gamma\delta/(1-\tau^2)$, generating dependence λ undistorted in both models to the graphs b), c); generating dependence α preserved with b), distorted with c).*



The equations of the summary graph model, resulting for Y_5 unobserved, have precisely one pair of correlated residuals, $\mathrm{cov}(\eta_1, \eta_3) = \gamma\delta$ and

$$Y_1 = \lambda Y_2 + \alpha Y_4 + \eta_1, \quad Y_2 = r_{23} Y_3 + \eta_2, \quad Y_3 = \tau Y_4 + \eta_3, \quad Y_4 = \eta_4.$$

The summary graph model preserves both λ and α as equation parameters.

In the corresponding maximal ancestral graph model, represented by the graph in Figure 2c), the equation parameters associated with arrows present in the graph are unconstrained linear least squares regression coefficients, see (2.19). These coefficients, expressed in terms of the generating parameters of

Figure 2a), are shown next to the arrows in Figure 2c). Thus, the generating coefficient λ is preserved, while α is changed into $\alpha - \tau\theta$, with $\theta = \gamma\delta/(1-\tau^2)$.

For generating graphs that are directed and acyclic, one can in general decide whether a distortion of a generating dependence, due to indirect confounding, may or may not occur in the parameters measuring the dependence in a maximal ancestral graph model by using graphical criteria for the corresponding summary graph, see Wermuth and Cox (2008). This paper extends such results to generating graphs that are multivariate regression chain graphs, see Section 5 below.

With some preliminary results in Section 2, we define in Section 3 a summary graph and some summary graph models. In Section 4, we obtain a summary graph from a summary graph in a larger node set. Our approach is based on matrix representations of graphs and on properties of matrix operators, see Wermuth, Wiedenbeck and Cox (2006), Wiedenbeck and Wermuth (2009). This method has been used by Marchetti and Wermuth (2009) to prove equivalence of different separation criteria for directed acyclic graphs. It is used here in Section 5 to interpret the matrix results in terms of paths.

2. Notation and preliminary results.

2.1. *Triangular systems of densities and the edge matrix of a parent graph.* For the stepwise process of generating the joint density f_V of a vector random variable of dimension $d_V \times 1$ in terms of univariate conditional densities $f_{i|i+1,\ldots,d_V}$, we start with the marginal density f_{d_V} of Y_{d_V}, proceed with the conditional density of Y_{d_V-1} given Y_{d_V}, up to Y_1 given $Y_2, \ldots, Y_{d_V}$.

A node k is named a parent node if an arrow starts at k and points to the offspring node i; the sets of such nodes are $\text{par}_i \subset \{i+1, \ldots, d_V\}$. We let $f_{i|\text{par}_i} = f_i$ whenever par_i is empty, assume that each of the univariate densities is non-degenerate and that

$$
\begin{aligned}
&(i) \;\; f_{i|i+1,\ldots,d_V} = f_{i|\text{par}_i} \text{ for each } i < d\,,\\
&(ii) \; f_{i|\text{par}_i} \neq f_{i|\text{par}_i \setminus l} \text{ for each } l \in \text{par}_i\,,\\
&(iii) \text{ specification of } f_{i|i+1,\ldots,d_V} \text{ not dependent on } f_{i+1,\ldots,d_V}\,.
\end{aligned}
\tag{2.1}
$$

Definition 1. *Any density f_V, obtained by this stepwise process and satisfying* (2.1) (i) *to* (iii)*, is said to be generated over a parent graph.*

Densities $f_{i|i+1,\ldots,d_V}$ obtained in this way show direct, non-vanishing dependences of variable Y_i precisely on the variables corresponding to parent nodes and on no others. Furthermore, the type and form of the conditional density of a variable generated in the future does not depend on the type and form of joint densities generated in the past. We now discuss consequences of each of the conditions in (2.1) in more detail since some subtle issues are involved.

Using node sets and the notation due to Dawid (1979), we write $a \perp\!\!\!\perp b|c$ for Y_a independent of Y_b given Y_c, where a, b, c are disjoint subsets of N. Condition (2.1) (i) is, for the above generating process of f_V, equivalently expressed in terms of conditional independence constraints

$$f_{i|i+1,\dots,d_V} = f_{i|\mathrm{par}_i} \iff i \perp\!\!\!\perp \{i+1,\dots,d_V\} \setminus k|\mathrm{par}_i \text{ for } k \notin \mathrm{par}_i\,. \tag{2.2}$$

With the sets of independence statements in (2.2), a graph in the fully ordered node set $V = (1, 2, \dots, d_V)$ is defined to have an ik-arrow, i.e. an edge that starts at node $k > i$ and points to node i, if and only if k is a parent node of i. We name it the parent graph, G^V_{par}. It is acyclic by construction.

The independence structure captured by G^V_{par} contains the statements (2.2) defining the graph and all those that can be derived from them, see e.g. Lemma 3 below. For Gaussian distributions generated over a given parent graph, there exist subfamilies which entail the independence structure captured by G^V_{par} but do not satisfy any additional constraints, see Geiger and Pearl (1993). We call them the relevant family of Gaussian distributions.

It is a consequence of condition (2.1)(i), that the joint density f_V factorizes as

$$f_V = \textstyle\prod_{i=1}^{d_V} f_{i|\mathrm{par}_i}. \tag{2.3}$$

Many authors start with a given set of local independence statements (2.2), which define a directed acyclic graph. They then consider joint distributions generated over directed acyclic graphs that are densities f_V that factorize as in (2.3). The class of distributions considered here is smaller, because the generating process permits only non-degenerate densities and because of conditions (2.1) $(ii), (iii)$.

Equation (ii) is a minimality condition. It excludes densities with any set of parent nodes smaller than par_i so that consequences of a specific model or of a specific family of models can be studied. Densities generated to satisfy (2.1) (i) and (ii) have been discussed as research hypotheses formulated in a given substantive context; see Wermuth and Lauritzen (1990). This notion is especially helpful whenever only dependences are to be considered that are strong enough to be of substantive interest while weak dependences are to be translated into conditional independence statements within the set (2.2).

Conditions (2.1) (ii) and (iii) together assure that a particular type of parametric cancellation does not occur since there is a unique independence statement associated with a missing edge for pair $\{i, j\}$ and edges present for both $\{i, k\}$ and $\{j, k\}$, either including or excluding the common neighbor node k. The contrary arises for an arbitrary conditioning set b, see Wermuth and Cox (2004), Section 7, if and only if

$$\int f_{ik|b} f_{jk|b} / f_{k|b} \, dy_k = f_{i|b} f_{j|b}, \tag{2.4}$$

a constraint connected to incomplete families of densities, i.e. those permitting zero expectation of a function that is not itself zero.

For any specific density generated over a parent graph, some types of parametric cancellations other than (2.4) may be still present. For instance, in a saturated, i.e. unconstrained, Gaussian distribution of three variables generated over a parent graph, it may hold that $1 \perp\!\!\!\perp 2$ even though Y_1 is dependent on Y_2 given Y_3. Often, such additional independences are judged to be of no substantive interest and avoidable in the corresponding family of distributions under study. This justifies the approach taken here, i.e. to study consequences of the independence structure captured by G^V_{par}, which in turn is defined via the above stepwise generating process.

Each graph that captures an independence structure has at least one binary matrix representations and a separate binary matrix for each type of edge. The edge matrix $\mathcal{A}$ of a parent graph is a $d_V \times d_V$ unit upper-triangular matrix, i.e. a matrix with ones along the diagonal and with zeros in the lower triangular part, such that for $i < k$, element $\mathcal{A}_{ik}$ satisfies

$$\mathcal{A}_{ik} = 1 \text{ if and only if } i \prec\!\!\!-\!\!\!- k \text{ in } G^V_{\text{par}}. \tag{2.5}$$

Because of the triangular form of the edge matrix $\mathcal{A}$ of G^V_{par}, a density f_V generated over a given parent graph, has also been called a triangular system of densities.

2.2. *Some more terminology for graphs.* A graph is defined by its node set and its edge sets, or equivalently, by its edge matrix components, one for each type of edge. If an edge is present in the graph for nodes i and k, then node pair $\{i, k\}$ is said to be coupled; otherwise it is said to be uncoupled.

An ik-path connects the path endpoint nodes i and k by a sequence of edges coupling distinct nodes. Nodes other than the endpoint nodes are the inner nodes of the path. If all inner nodes in a path are in set a, then the path is called an a-line path. An edge is regarded as a path without inner nodes. For a graph in node set N and $a \subset N$, the subgraph induced by a is obtained by removing all nodes and edges outside a.

Both a graph and a path are called directed if all its edges are arrows. If in a directed path an arrow starts at node k and all arrows of the path point in the direction of node i, then node k is an ancestor of i, node i a descendant of k, and the ik-path is called a descendant-ancestor path.

2.3. *Linear triangular systems.* For a parent graph with edge matrix (2.5), a linear triangular system is given by a set of recursive linear equations for a mean-centred random vector variable Y of dimension $d_V \times 1$ having $\text{cov}(Y) = \Sigma$, i.e. by

$$AY = \varepsilon, \tag{2.6}$$

where A is a real-valued $d_V \times d_V$ unit upper-triangular matrix, given by

$$A_{ik} = 0 \iff \mathcal{A}_{ik} = 0, \qquad E_{\text{lin}}(Y_i | Y_{i+1} = y_{i+1}, \ldots, Y_d = y_d) = -A_{i,\text{par}_i} y_{\text{par}_i},$$

and $E_{\text{lin}}(\cdot)$ denotes a linear predictor, see e.g. (2.19). The random vector ε of residuals has zero mean and $\text{cov}(\varepsilon) = \Delta$, a diagonal matrix with $\Delta_{ii} > 0$. A Gaussian triangular system is generated if the distribution of each residual ε_i is Gaussian.

The covariance and concentration matrix of Y are, respectively,

$$\Sigma = A^{-1}\Delta(A^{-1})^{\mathrm{T}}, \qquad \Sigma^{-1} = A^{\mathrm{T}}\Delta^{-1}A. \tag{2.7}$$

Thus, the linear independences that constrain the equations (2.6) are defined by zeros in the triangular decomposition, (A, Δ^{-1}), of the concentration matrix. The edge matrix $\mathcal{A}$ of G^V_{par} coincides for linear triangular systems generated over G^V_{par} with the indicator matrix of zeros in A, i.e. $\mathcal{A} = \text{In}[A]$, where $\text{In}[\cdot]$ changes every nonzero entry of a matrix into a one. For the relevant family of Gaussian distribution, the list of pairs with only non-vanishing dependences are given by the set of ij-ones in $\mathcal{A}$ for $i < j$.

It is a property only of joint Gaussian distributions of Y that probabilistic and linear independence statements coincide, but for every density generated over G^V_{par}, probabilistic independence statements combine just like linear independences, see Lemma 1 of Marchetti and Wermuth (2009). Therefore, transformations of the edge matrix $\mathcal{A}$, that mimic linear transformations of A, are useful for studying consequences of parent graphs in general.

Edge matrices expressed in terms of components of set of given generating edge matrices are called induced. Examples of edge matrices induced by $\mathcal{A}$ are the overall covariance and the overall concentration graph, see Wermuth and Cox (2004). These edge matrices are, respectively,

$$\mathcal{S}_{VV} = \text{In}[\mathcal{A}^-(\mathcal{A}^-)^{\mathrm{T}}], \quad \mathcal{S}^{VV} = \text{In}[\mathcal{A}^{\mathrm{T}}\mathcal{A}],$$

where $\mathcal{A}^- = \mathcal{A}$ but $\mathcal{A}^-$ having an additional one compared to $\mathcal{A}$ in position (i, k) if and only if k is an ancestor but not a parent of i in G^V_{par}.

Both matrices are symmetric and mimic the expressions for the covariance and the concentration matrix implied by a linear triangular system, given in (2.7). For triangular systems of densities, a zero in position (i,k) of $\mathcal{S}_{VV}$ and of $\mathcal{S}^{VV}$ means, respectively, that

$$i \perp\!\!\!\perp k, \quad i \perp\!\!\!\perp k | V \setminus \{i,k\}$$

is implied for every density generated over a given parent graph that has edge matrix $\mathcal{A}$. More complex induced edge matrices are derived in the following.

2.4. *Partial inversion and partial closure.* Let F be a square matrix of dimension d_V with principal submatrices that are all invertible and $\mathcal{F}$ be an associated binary edge matrix in node set $V = \{1, \ldots, d_V\}$.

The operator called partial closure, applied to edge set V, transforms $\mathcal{F}$ into $\text{zer}_V \mathcal{F} = \mathcal{F}^-$, the edge matrix of a graph in which all paths of special type are closed, see here Section 5. The operator called partial inversion, applied to the index set V transforms F into its inverse, $\text{inv}_V F = F^{-1}$. When applying the operators to an arbitrary subset a of V, the just described overall operations are modified into closing only a-line paths and to inverting matrices only partially; see Wermuth, Wiedenbeck and Cox (2006), Section 2, for proofs and discussions of the results in equations (2.11) to (2.14) below.

Let F and $\mathcal{F}$ be partitioned in the order (a,b). The effect of applying partial closure (2.9) to rows and columns a of the edge matrix $\mathcal{A}$ of a parent graph, i.e. to rows and columns of $\mathcal{A}$, is to keep all arrows present and to add arrows by turning every a-line ancestor into a parent. By applying partial inversion to a of F, the linear equations $FY = \eta$, say, are modified into

$$\text{inv}_a F \begin{pmatrix} \eta_a \\ Y_b \end{pmatrix} = \begin{pmatrix} Y_a \\ \eta_b \end{pmatrix}. \tag{2.8}$$

DEFINITION 2. Matrix formulations of $\text{inv}_a F$, $\text{zer}_a \mathcal{F}$. *In explicit form*

$$\text{inv}_a F = \begin{pmatrix} F_{aa}^{-1} & -F_{aa}^{-1} F_{ab} \\ F_{ba} F_{aa}^{-1} & F_{bb.a} \end{pmatrix}, \quad \text{zer}_a \mathcal{F} = \text{In}[\begin{pmatrix} \mathcal{F}_{aa}^- & \mathcal{F}_{aa}^- \mathcal{F}_{ab} \\ \mathcal{F}_{ba} \mathcal{F}_{aa}^- & \mathcal{F}_{bb.a} \end{pmatrix}], \tag{2.9}$$

$$F_{bb.a} = F_{bb} - F_{ba} F_{aa}^{-1} F_{ab}, \qquad \mathcal{F}_{bb.a} = \text{In}[\mathcal{F}_{bb} + \mathcal{F}_{ba} \mathcal{F}_{aa}^- \mathcal{F}_{ab}],$$

and

$$\mathcal{F}_{aa}^- = \text{In}[(n \mathcal{I}_{aa} - \mathcal{F}_{aa})^{-1}], \tag{2.10}$$

where $n - 1 = d_a$ denotes the dimension of $\mathcal{F}_{aa}$ and $\mathcal{I}_{aa}$ is an identity matrix of dimension d_a.

The inverse in (2.10) has a zero entry if and only if there is a structural zero in F_{aa}^{-1}, i.e. a zero that is preserved for all permissible values in F_{aa}. For instance with $F_{aa} = A$ of (2.6), the permissible values are those that lead to a positive definite concentration matrix $A^{\mathrm{T}}\Delta^{-1}A$.

Note that we have $\mathrm{zer}_a\mathcal{F} = \mathrm{In}[\mathrm{inv}_a F]$ if and only if there is no zero in $\mathrm{inv}_a F$ caused by parametric cancellation. Otherwise, there may be additional zeros in $\mathrm{In}[\mathrm{inv}_a F]$ compared to the edge matrix $\mathrm{zer}_a\mathcal{F}$.

It follows directly from (2.8) that F partially inverted on a coincides with F^{-1} partially inverted on $V \setminus a$

$$\mathrm{inv}_a F = \mathrm{inv}_{V\setminus a} F^{-1}\,. \tag{2.11}$$

Some further properties of the operators are needed here later. Both operators are commutative so that, for $V = \{a, b, c, d\}$,

$$\mathrm{inv}_a\,\mathrm{inv}_b F = \mathrm{inv}_b\,\mathrm{inv}_a F, \qquad \mathrm{zer}_a\,\mathrm{zer}_b\,\mathcal{F} = \mathrm{zer}_b\,\mathrm{zer}_a\mathcal{F}, \tag{2.12}$$

and both operations can be exchanged with selecting a submatrix so that, for $J = \{a, b\}$,

$$[\mathrm{inv}_a F]_{J,J} = \mathrm{inv}_a F_{JJ}, \qquad [\mathrm{zer}_a\mathcal{F}]_{J,J} = \mathrm{zer}_a\mathcal{F}_{JJ}, \tag{2.13}$$

but partial inversion can be undone while partial closure cannot

$$\mathrm{inv}_{ab}\,\mathrm{inv}_{bc} F = \mathrm{inv}_{ac} F, \qquad \mathrm{zer}_{ab}\,\mathrm{zer}_{\,bc}\,\mathcal{F} = \mathrm{zer}_{abc}\,\mathcal{F}. \tag{2.14}$$

EXAMPLE 1. Partial inversion applied to Σ and to Σ^{-1}. *The symmetric covariance matrix* Σ *and the concentration matrix* Σ^{-1} *of* Y *are written, partitioned according to* (a, b)*, as*

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ . & \Sigma_{bb} \end{pmatrix}, \qquad \Sigma^{-1} = \begin{pmatrix} \Sigma^{aa} & \Sigma^{ab} \\ . & \Sigma^{bb} \end{pmatrix},$$

where the . notation indicates the symmetric entry. Partial inversion of Σ^{-1} *on* a *gives* $\Pi_{a|b}$*, the population coefficient matrix of* Y_b *in linear least squares regression of* Y_a *on* Y_b*, defined by*

$$E_{\mathrm{lin}}(Y_a|Y_b = y_b) = \Pi_{a|b}\, y_b,$$

the covariance matrix $\Sigma_{aa|b}$ *of* $Y_{a|b} = Y_a - \Pi_{a|b}Y_b$ *and the marginal concentration matrix* $\Sigma^{bb.a}$ *of* Y_b

$$\mathrm{inv}_a\Sigma^{-1} = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma^{bb.a} \end{pmatrix}, \tag{2.15}$$

where the $\sim$ notation denotes entries in a matrix which is symmetric except for the sign. Property (2.11), $\mathrm{inv}_a\Sigma^{-1} = \mathrm{inv}_b\Sigma$, *leads at once to several well known dual expressions for the three submatrices in* (2.15), *by writing the two partial inversions explicitly*

$$\begin{pmatrix} (\Sigma^{aa})^{-1} & -(\Sigma^{aa})^{-1}\Sigma^{ab} \\ \sim & \Sigma^{bb} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ab} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} & \Sigma_{ab}\Sigma_{bb}^{-1} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix}.$$

Variants of $(\Sigma^{aa})^{-1} = \Sigma_{aa|b}$ and of $\Sigma_{bb}^{-1} = \Sigma^{bb.a}$ will be used repeatedly.

2.5. *The operators applied to block-triangular systems.* For a system of linear equations, in a mean-centred vector variable Y having covariance matrix Σ, that is block-triangular in two ordered blocks (a, b) with

$$HY = \eta, \text{ with } H_{ba} = 0, \quad \mathrm{cov}(\eta) = W \text{ positive definite}, \tag{2.16}$$

the concentration matrix $H^{\mathrm{T}}W^{-1}H$ can be partially inverted by combining partially inverted components of H and W^{-1}.

For this result, obtained by direct computation or by use of Theorem 1 in Wermuth and Cox (2004), we let

$$K = \mathrm{inv}_a H, \qquad Q = \mathrm{inv}_b W.$$

LEMMA 1. Partially inverted matrix product $H^{\mathrm{T}}W^{-1}H$ for H block-triangular in (a, b).

$$\mathrm{inv}_a(H^{\mathrm{T}}W^{-1}H) = \begin{pmatrix} K_{aa}Q_{aa}K_{aa}^{\mathrm{T}} & K_{ab} + K_{aa}Q_{ab}K_{bb} \\ \sim & H_{bb}^{\mathrm{T}}Q_{bb}H_{bb} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix}. \tag{2.17}$$

To obtain the edge matrix of the three components of $\mathrm{inv}_a\Sigma^{-1}$ induced by a parent graph, an additional argument is needed. We assume that the edge matrices for H, W, are given by $\mathcal{H}$, $\mathcal{W}$ and let $\mathcal{K} = \mathrm{zer}_a\mathcal{H}$ and $\mathcal{Q} = \mathrm{zer}_b\mathcal{W}$.

LEMMA 2. The edge matrix components induced for $\mathrm{inv}_a\Sigma^{-1}$ by the edge matrices $\mathcal{H}, \mathcal{W}$ of the block-triangular system (2.16). *Structural zeros of* $\mathrm{inv}_a\Sigma^{-1}$ *are given by zeros in*

$$\begin{pmatrix} \mathcal{S}_{aa|b} & \mathcal{P}_{a|b} \\ . & \mathcal{S}^{bb.a} \end{pmatrix} = \mathrm{In}[\begin{pmatrix} \mathcal{K}_{aa}\mathcal{Q}_{aa}\mathcal{K}_{aa}^{\mathrm{T}} & \mathcal{K}_{ab} + \mathcal{K}_{aa}\mathcal{Q}_{ab}\mathcal{K}_{bb} \\ . & \mathcal{H}_{bb}^{\mathrm{T}}\mathcal{Q}_{bb}\mathcal{H}_{bb} \end{pmatrix}], \tag{2.18}$$

where the induced edge matrix components are, respectively,

$$\mathcal{S}_{aa|b}, \ \mathcal{P}_{a|b}, \ \mathcal{S}^{bb.a} \text{ for } \Sigma_{aa|b}, \ \Pi_{a|b}, \ \Sigma^{bb.a} = \Sigma_{bb}^{-1}.$$

Proof. The submatrices of H, K and Q in $\mathrm{inv}_a\Sigma^{-1} = H^{\mathrm{T}}W^{-1}H$ of Lemma 1 are expressed without any self-canceling matrix operations such as a matrix multiplied by its inverse. When these are replaced by submatrices of the non-negative matrices $\mathcal{H}, \mathcal{K}$ and $\mathcal{Q}$, which have the appropriate structural zeros, then just the structural zeros in $\mathrm{inv}_a\Sigma^{-1}$ are preserved by multiplying, summing and applying the indicator function. $\square$

The graphs with edge matrices $\mathcal{S}_{aa|b}$ and $\mathcal{S}^{bb.a}$ have been named the conditional covariance graph of Y_a given Y_b and the marginal concentration graph of Y_b (Wermuth and Cox, 1998) in which a missing ik-edge represents, respectively,

$$i \perp\!\!\!\perp k|b \text{ and } i \perp\!\!\!\perp k|b \setminus \{i,k\}.$$

The rectangular edge matrix $\mathcal{P}_{a|b}$ represents conditional dependence of Y_i, $i \in a$, on Y_j, $j \in b$, given $Y_{b\setminus j}$ so that a missing ik-edge means $i \perp\!\!\!\perp k|b \setminus k$.

2.6. *Generating constrained linear multivariate regressions.* Next, we describe two special ways of generating equations in $Y_{a|b}$ and Y_b and associated edge matrix results that will be used below for interpreting graphs.

2.6.1. *Some notation for multivariate regressions and some known results.* The equations for a linear multivariate regression model in mean-centred variables with covariance matrix Σ and regressing Y_a on Y_b, may be written as

$$Y_a = \Pi_{a|b}Y_b + \epsilon_a, \text{ with } \mathrm{E}(\epsilon_a) = 0, \ \mathrm{cov}(\epsilon_a, Y_b) = 0. \tag{2.19}$$

For residuals ϵ_a uncorrelated with the regressor variables Y_b, taking expectations in the equation given by $Y_aY_b^{\mathrm{T}}$ defines $\Pi_{a|b}$ via $\Sigma_{ab} = \Pi_{a|b}\Sigma_{bb}$ and $\Sigma_{aa|b} = \mathrm{cov}(Y_a - \Pi_{a|b}Y_b) = \mathrm{cov}(\epsilon_a)$.

We write the population matrix of the least-squares regression coefficients $\Pi_{a|b}$ partitioned according to (β, γ) with $\beta \subset b$ and $\gamma = b \setminus \beta$ as

$$\Pi_{a|b} = \begin{pmatrix} \Pi_{a|\beta.\gamma} & \Pi_{a|\gamma.\beta} \end{pmatrix},$$

and note that, for example, $\Pi_{a|\beta.\gamma}$ is both the coefficient matrix of Y_β in model (2.19) and the coefficient matrix of $Y_{\beta|\gamma}$ in linear least-squares regression of $Y_{a|\gamma}$ on $Y_{\beta|\gamma}$ that is after both Y_a and Y_β are adjusted for linear dependence on Y_γ, i.e. in

$$Y_{a|\gamma} = \Pi_{a|\beta.\gamma}Y_{\beta|\gamma} + \epsilon_a, \text{ with } \mathrm{E}(\epsilon_a) = 0, \ \mathrm{cov}(\epsilon_a, Y_{\beta|\gamma}) = 0, \tag{2.20}$$

with residuals ϵ_a unchanged compared to model (2.19).

This may be proven using Example 1. By moving from $\mathrm{inv}_\gamma\Sigma$ to $\mathrm{inv}_b\Sigma = \mathrm{inv}_\beta(\mathrm{inv}_\gamma\Sigma)$, first the parameter matrices for both of the equations $Y_{a|\gamma} = Y_a - \Pi_{a|\gamma}Y_\gamma$ and $Y_{\beta|\gamma} = Y_\beta - \Pi_{\beta|\gamma}Y_\gamma$ are obtained, then we get for $Y_{a|b}$ in the second step $\Pi_{a|\beta.\gamma} = \Sigma_{a\beta|\gamma}\Sigma^{-1}_{\beta\beta|\gamma}$ and

$$\Pi_{a|\gamma.\beta} = \Pi_{a|\gamma} - \Pi_{a|\beta.\gamma}\Pi_{\beta|\gamma}. \tag{2.21}$$

Equation (2.21) is known as the matrix form of Cochran's recursive relation among regression coefficients. It leads, for instance, to conditions under which a marginal and a partial regression coefficient matrix coincide. Given edge matrices $\mathcal{P}_{a|b}$ and $\mathcal{P}_{a|\gamma}$ induced by a parent graph, equation (2.21) implies, for all distributions generated over a parent graph

$$\mathcal{P}_{a|\gamma.\beta} = \mathcal{P}_{a|\gamma} \text{ if } \mathcal{P}_{a|\beta.\gamma} = 0. \tag{2.22}$$

2.6.2. *Two ways of generating $Y_{a|b}$ and constraints on $\Pi_{a|b}$.* Suppose a linear system of equations is block-triangular in $N = (a, b)$, then it is also orthogonal in (a, b) if to the equations (2.16), the condition $\mathrm{cov}(\eta_a, \eta_b) = 0$ is added so that $0 = W_{ab} = W^{\mathrm{T}}_{ba}$.

After partial inversion on a, the linear multivariate regression model (2.19) with

$$\Pi_{a|b} = -H^{-1}_{aa}(H_{a\beta} \quad H_{a\gamma}), \tag{2.23}$$

results, see equations (2.17). In econometrics, such models have been named reduced form equations. An analogue of Cochran's equation (2.21) is then

$$-H_{aa}\Pi_{a|\gamma} = H_{a\gamma} + H_{a\beta}\Pi_{\beta|\gamma}. \tag{2.24}$$

For b split as before and $\alpha \subset a$ and $\delta = a \setminus \alpha$, the matrix identity

$$\Pi_{\alpha|\beta.\gamma} = [\Pi_{a|b}]_{\alpha,\beta} = -H^{-1}_{\alpha\alpha.\delta}H_{\alpha\beta.\delta}$$

gives the coefficient of $Y_{\beta|\gamma}$ in linear least-squares regression of $Y_{\alpha|\gamma}$ on $Y_{\beta|\gamma}$. Thus for such order-compatible splits, in which $\alpha \subset a$ and $\beta \subset b$, all densities generated over parent graphs and having induced edge matrices $\mathcal{H}$ and $\mathcal{W}$ such that $\mathcal{H}_{ba} = \mathcal{W}_{ba} = 0$ and $\mathcal{W}_{ab} = 0$, satisfy

$$\alpha \perp\!\!\!\perp \beta|\gamma \iff \mathcal{P}_{\alpha|\beta.\gamma} = 0 \iff \mathcal{H}_{\alpha\beta.\delta} = 0. \tag{2.25}$$

When we start instead with a mean-centred random vector Y and with zero constraints on equation parameters in the following concentration equations,

$$\Sigma^{-1}Y = \zeta \text{ with } \mathrm{cov}(Y) = \Sigma, \tag{2.26}$$

then the equation parameters coincide with the residual covariance matrix, $\mathrm{cov}(\zeta) = \Sigma^{-1}$.

The relations after partial inversion on a are

$$\mathrm{inv}_a \Sigma^{-1} \begin{pmatrix} \zeta_a \\ Y_b \end{pmatrix} = \begin{pmatrix} Y_a \\ \zeta_b \end{pmatrix}.$$

These give constrained orthogonal equations in $Y_{a|b}$ and in Y_b, with $\Pi_{a|b} = -(\Sigma^{aa})^{-1}\Sigma^{ab}$, and

$$Y_a = \Pi_{a|b} Y_b + \Sigma_{aa|b}\zeta_a, \qquad \Sigma^{bb.a} Y_b = \zeta_{b|a}, \qquad \zeta_{b|a} = \zeta_b + \Pi^{\mathrm{T}}_{a|b}\zeta_a, \tag{2.27}$$

where $\mathrm{cov}(\zeta_a) = \Sigma^{aa}$, $\mathrm{cov}(\zeta_{b|a}) = \Sigma^{bb.a} = \Sigma^{-1}_{bb}$ and $\mathrm{cov}(\zeta_a, \zeta_{b|a}) = 0$.

Thus, for densities in which the independence structure is captured by the concentration graph of Y_V, it holds for $V = \{\alpha, \beta, \gamma, \delta\}$ that

$$\alpha \perp\!\!\!\perp \beta | \gamma \iff \mathcal{S}^{\alpha\beta.\delta} = 0. \tag{2.28}$$

3. Summary graphs and associated models.

As we shall see, summary graphs have local and global Markov properties and Gaussian summary graph models are special block-triangular equations (2.16).

3.1 Definitions and properties of Gaussian summary graph models. As is to be described in more detail in Sections 3.3 and 3.4, starting from a Gaussian triangular system (2.6) for Y_V with $V = \{N, C, M\}$, conditioning on Y_C and marginalising over Y_M defines remaining variables Y_v in the past of Y_C, remaining variables Y_u not in the past of Y_C and equations in $Y_{N|C}$ for $N = (u, v)$ of the following form.

DEFINITION 3. Gaussian summary graph model. *A Gaussian summary graph model is given by a system of equations $HY_{N|C} = \eta$, that is block-triangular and orthogonal in (u, v) with*

$$\begin{pmatrix} H_{uu} & H_{uv} \\ 0 & \Sigma^{-1}_{vv|C} \end{pmatrix} \begin{pmatrix} Y_{u|C} \\ Y_{v|C} \end{pmatrix} = \begin{pmatrix} \eta_u \\ \zeta_v \end{pmatrix}, \qquad \mathrm{cov}\begin{pmatrix} \eta_u \\ \zeta_v \end{pmatrix} = \begin{pmatrix} W_{uu} & 0 \\ . & \Sigma^{-1}_{vv|C} \end{pmatrix}, \tag{3.1}$$

where H_{uu} is unit upper-triangular, W_{uu} and $\Sigma_{vv|C}$ are symmetric, and each of η_u and ζ_v have non-degenerate Gaussian distributions. Zero constraints on W_{uu}, $\Sigma^{-1}_{vv|C}$ and additional zero constraints on H are captured by the graph G^N_{sum}, defined below.

Equation (3.1) implies for $Y_{u|C}$ given $Y_{v|C}$ a constrained multivariate, Gaussian regression model (2.19) and for $Y_{v|C}$ a Gaussian concentration graph model. The latter had been introduced under the name of covariance selection by Dempster (1972). The residuals of $Y_{u|C}$ and $Y_{v|C}$ are uncorrelated so that by (2.23), $\Pi_{u|v.C} = -H_{uu}^{-1}H_{uv}$ and by (2.20), the equations in $Y_{u|C}$ of (3.1) can also be written as

$$H_{uu}Y_{u|vC} = \eta_u \text{ with } W_{uu} = H_{uu}\Sigma_{uu|vC}H_{uu}^{\mathrm{T}}. \tag{3.2}$$

These specify what in econometrics has been called a recursive system of regressions in endogenous variables $Y_{u|vC}$. The equation parameter matrix H_{uu} is, as in the linear triangular system (2.6), of unit upper-triangular form, but by contrast to a triangular system, the residuals η_u are in general correlated.

Identification is an issue for the equation parameters H_{uu} in (3.2), but it is assured under the general condition that for any pair (i,k) within u either $H_{ik} = 0$ or $W_{ik} = 0$, see e.g. Brito and Pearl (2002). This shows in a summary by the absence of any double edge. The fit of any under-identified model (3.1) to data may be judged with the help of the independence equivalent, maximal ancestral graph model obtained in equation (3.6) below.

Gaussian multivariate regression chains arise as an important special case of (3.2). A partitioning of the node set $(1, 2, \dots, d_u)$ into chain components, g_j, defines with the ordered sequence $N \setminus v = (g_1, \dots, g_j, \dots, g_J)$ and sets $r_j = \{g_{j+1}, ..., g_J\}$ a block-recursive factorization of the joint conditional density $f_{u|vC}$ as

$$f_{u|vC} = \textstyle\prod_{j=1}^{J} f_{g_j|r_j}\,, \tag{3.3}$$

and Gaussian multivariate regressions (2.19) of Y_{g_j} on Y_{r_j}. A zero ik-element in the residual covariance matrix and in the regression coefficient matrix, means respectively,

$$i \perp\!\!\!\perp k | r_j \text{ for } i,k \in g_j, \qquad i \perp\!\!\!\perp k | r_j \setminus \{k\} \text{ for } i \in g_j \text{ and } k \in r_j. \tag{3.4}$$

3.2 Summary graphs and their local Markov properties. We now present graphs denoted by G^N_{sum}. We name them summary graphs since they will be shown to summarize precisely those independences implied by a parent graph G^V_{par} for Y_N conditioned on Y_C, where $N = V \setminus \{C, M\}$.

DEFINITION 4. Summary graph. *A summary graph, G^N_{sum}, has node set N and the following edge matrix components, where each component is a binary matrix and each square matrix has ones along the diagonal,*

$\mathcal{H}_{uu}$, upper-triangular, and $\mathcal{H}_{uv}$, rectangular, both for arrows pointing to u,

$$\begin{aligned} (3.5) \qquad & \mathcal{W}_{uu}, \ \textit{symmetric, for dashed lines within } u, \\ & \mathcal{S}^{vv.uM}, \ \textit{symmetric, for full lines within } v. \end{aligned}$$

For $i < k$, there is an ik-zero in one of the edge matrix components if and only if the corresponding edge is missing in G^N_{sum}. For a Gaussian block-triangular system (3.1)*, the edge matrix components $\mathcal{H}_{uu}, \mathcal{H}_{uv}, \mathcal{W}_{uu}$, and $\mathcal{S}^{vv.uM}$ of a summary graph are for the parameter matrices H_{uu}, H_{uv}, W_{uu}, and $\Sigma^{vv.uM} = \Sigma^{-1}_{vv|C}$.*

To derive local Markov properties of G^N_{sum}, we note first that the i'th equation in (3.2) is modified by an orthogonalising step into a linear least-squares regression equation of $Y_{i|vC}$ on $Y_{d|vC}$, when d is the set of ancestor nodes of i in u. This gives via least-squares regression of residual η_i on η_d

$$(3.6) \quad -\Pi_{i|d.Cv} = H_{id} - W_{id}W^{-1}_{dd}H_{dd} \text{ and } \mathcal{P}_{i|d.Cv} = \text{In}[\mathcal{H}_{id} + \mathcal{W}_{id}\mathcal{W}^{-}_{dd}\mathcal{H}_{dd}].$$

Thus, for an uncoupled node pair (i, k) with $k \in d$, no independence statement is implied for Y_i, Y_k if $\mathcal{W}_{id}\mathcal{W}^{-}_{dd}\mathcal{H}_{dk} \neq 0$. Also, $W_{id}W^{-1}_{dd}H_{dd}$ quantifies the distortion introduced in the least-squares regression coefficient vector $\Pi_{i|d.Cv}$ for the vector of equation parameters H_{id}. The former are the equation parameters in Gaussian maximal ancestral graph models.

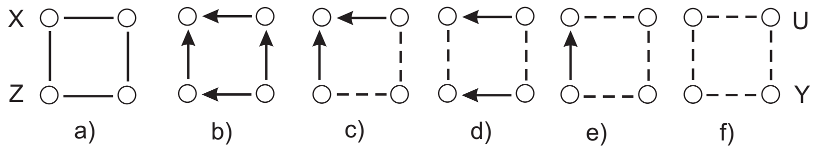
PROPOSITION 1. Local Markov properties of summary graphs. *Let β denote subsets of N uncoupled to node i in G^N_{sum} of* (3.5) *which has edge matrix components $\mathcal{H}_{uN}$, $\mathcal{W}_{uu}$ and $\mathcal{S}^{vv.uM}$. Let further d contain all ancestor nodes of i in u, then the local Markov properties of G^N_{sum} are*
(i) for $i \in u$ and $\beta \subset d$: $i \perp\!\!\!\perp \beta | Cvd \setminus \beta \iff \mathcal{H}_{i\beta} = 0$ and $\mathcal{W}_{id}\mathcal{W}^{-}_{dd}\mathcal{H}_{d\beta} = 0$,
(ii) for $i \in u$ and $\beta \subset v$: $i \perp\!\!\!\perp \beta | Cv \setminus \beta \iff \mathcal{H}_{i\beta.d} = 0$,
(iii) for $i \in v$ and $\beta \subset v$: $i \perp\!\!\!\perp \beta | Cv \setminus \{i, \beta\} \iff [\mathcal{S}^{vv.uM}]_{i,\beta} = 0$.

PROOF. The edge matrix condition (i) results with equations (3.6), (ii) with (2.25) and (iii) with (2.28). □

The following figure shows special cases of summary graphs, noting that C and one of u, v may be empty sets.

It also shows that summary graphs cover all six possible combinations of independence constraints on two non-overlapping pairs of the four variables X, Z, U, Y. Substantive research examples with well-fitting data to linear models of Figure 3 have been given by Cox and Wermuth (1993) to the concentration graph in a), the parent graph in b), the graph of seemingly unrelated regressions in d) and the covariance graph in f).

FIG 3. *Important special cases. Two non-overlapping pairs are constrained:* X, Y *and* Z, U*; with* $X \perp\!\!\!\perp Y | ZU$ *in a),b),c), with* $X \perp\!\!\!\perp Y | U$ *in d),e), and with* $X \perp\!\!\!\perp Y$ *in e); with* $Z \perp\!\!\!\perp U$ *in c),e), f), with* $Z \perp\!\!\!\perp U | Y$ *in b),d) and with* $Z \perp\!\!\!\perp U | XY$ *in a).*



3.3 Generating G^N_{sum} *from* G^V_{par}. Starting from a Gauusian triangular system in (2.6) with parent graph G^V_{par}, the choice of any conditioning set C leads to an ordered split $V = (O, R)$, where we think of $R = \{C, F\}$ as the nodes to the right of C, see equation (3.7). Every node in F is an ancestor of a node in C, so that we call F the set of foster nodes of C. No node in O has a descendant in R so that O is said to contain the outsiders of R. Equations, orthogonal and block-triangular in (O, R), are

$$\begin{pmatrix} A_{OO} & A_{OR} \\ 0 & A_{RR} \end{pmatrix} \begin{pmatrix} Y_O \\ Y_R \end{pmatrix} = \begin{pmatrix} \varepsilon_O \\ \varepsilon_R \end{pmatrix}. \tag{3.7}$$

After conditioning on Y_C and marginalising over Y_M, the resulting system preserves block-triangularity and orthogonality with $u \subset O$, $v \subset F$.

PROPOSITION 2. Equations and graph obtained after conditioning on Y_C, then marginalising over Y_M. *Given a Gaussian triangular system* (2.6) *generated over* G^V_{par}, *conditioning set* C, *marginalising set* $M = (p, q)$ *with*

$$p = O \setminus u, \qquad q = F \setminus v,$$

and parameter and edge matrices, arranged in the appropriate order,

$$D = \text{inv}_p \tilde{A}, \qquad \mathcal{D} = \text{zer}_p \tilde{\mathcal{A}},$$

$$\Sigma^{FF.O} = [A^{\mathrm{T}}_{RR} \Delta^{-1}_{RR} A_{RR}]_{F,F}, \qquad \mathcal{S}^{FF.O} = \text{In}[\mathcal{A}^{\mathrm{T}}_{RR} \mathcal{A}_{RR}]_{F,F},$$

$$\text{inv}_q \Sigma^{FF.O} = \begin{pmatrix} \Sigma_{qq|vC} & \Pi_{q|v.C} \\ \sim & \Sigma^{-1}_{vv.C} \end{pmatrix}, \qquad \text{zer}_q\, \mathcal{S}^{FF.O} = \begin{pmatrix} \mathcal{S}_{qq|vC} & \mathcal{P}_{q|v.C} \\ . & \mathcal{S}^{vv.qO} \end{pmatrix}.$$

After first conditioning on Y_C *and removing* Y_C, *then marginalising over* Y_M *and removing* Y_M, *the induced linear equations* (3.1) *in* $Y_{N|C}$ *have*

$$H_{uu} = D_{uu}, \quad H_{uv} = D_{uv} + D_{uq} \Pi_{q|v.C}, \tag{3.8}$$

$$W_{uu} = (\Delta_{uu} + D_{up} \Delta_{pp} D^{\mathrm{T}}_{up}) + (D_{uq} \Sigma_{qq|vC} D^{\mathrm{T}}_{uq}); \tag{3.9}$$

the induced edge matrix components of the summary graph $G^N_{\rm sum}$ are

$$\mathcal{H}_{uu} = \mathcal{D}_{uu}, \quad \mathcal{H}_{uv} = \text{In}[\mathcal{D}_{uv} + \mathcal{D}_{uq}\mathcal{P}_{q|v.C}], \tag{3.10}$$

$$\mathcal{W}_{uu} = \text{In}[(\mathcal{I}_{uu} + \mathcal{D}_{up}\mathcal{D}^{\rm T}_{up}) + (\mathcal{D}_{uq}\mathcal{S}_{qq|vC}\mathcal{D}^{\rm T}_{uq})]. \tag{3.11}$$

Proof. From the equations (3.7) in Y, the equations in $Y_{O|C}$ and $Y_{F|C}$

$$A_{OO}Y_{O|C} + A_{OF}Y_{F|C} = \varepsilon_O, \qquad \Sigma^{-1}_{FF|C}Y_{F|C} = \zeta_F, \qquad \zeta_R = A^{\rm T}_{RR}\Delta^{-1}_{RR}\varepsilon_R\,,$$

are obtained by using (2.20), (2.26) and (2.27). Then, equations (3.1) in $Y_{u|C}$, $Y_{v|C}$ result, with parameters given in (3.8), (3.9), after partial inversion on $M = (p, q)$ and deleting the equations in $Y_{M|C}$. Thereby is $p \subset O$, $q \subset F$ and

$$\text{inv}_M\begin{pmatrix} \tilde{A}_{OO} & \tilde{A}_{OF} \\ 0 & \tilde{\Sigma}^{-1}_{FF|C}\end{pmatrix}\begin{pmatrix}\varepsilon_p \\ Y_{u|C} \\ \zeta'_q \\ Y_{v|C}\end{pmatrix} = \begin{pmatrix} Y_{p|C} \\ \varepsilon_u \\ Y_{q|C} \\ \zeta'_v\end{pmatrix}.$$

The uncorrelated residuals are

$$\eta_u = (\varepsilon_u - D_{up}\varepsilon_p) - D_{uq}\Sigma_{qq|vC}\zeta_q, \qquad \zeta_v = \zeta'_v + \Pi^{\rm T}_{q|v.C}\zeta'_q.$$

After replacing the defining matrix components in (3.8), (3.9) by their corresponding edge matrix components and applying the indicator function, the induced edge matrix components (3.10), (3.11) of $G^N_{\rm sum}$ are obtained. □

It is instructive to also check the relations of the parameter matrices in (3.8), (3.9) to multivariate regression coefficients and to conditional covariance matrices. With $\Pi_{u|R} = -D^{-1}_{uu}(D_{uv},\ D_{uq},\ D_{uC})$, equation (2.24) gives

$$-D_{uu}\Pi_{u|v.C} = D_{uv} + D_{uq}\Pi_{q|v.C}, \qquad D_{uu}(Y_{u|C} - \Pi_{u|v.C}Y_{v|C}) = D_{uu}Y_{u|vC},$$

and for W_{uu} defined in (3.2) and specialized in (3.9)

$$D^{-1}_{uu}W_{uu}D^{-T}_{uu} = \Sigma_{uu|vqC} + \Pi_{u|q.vC}\Sigma_{qq|vC}\Pi^{\rm T}_{u|q.vC} = \Sigma_{uu|vC},$$

so that the required relations are obtained for $Y_{u|vC}$ and $Y_{v|C}$.

3.5 Models associated with summary graphs. As noted before, the density $f_{N|C}$ of Y_N given Y_C is well-defined since it is obtained from a density of Y_V generated over a parent graph by marginalising over Y_M and conditioning on Y_C. As we have seen, this leads to the factorization of $f_{N|C}$ into $f_{u|vC}$ and $f_{v|C}$. The independence structure of Y_v given Y_C is captured by a concentration graph. Corresponding models have for instance been studied by Lauritzen and Wermuth (1989), extending the Gaussian covariance selection models and the graphical, log-linear interaction models for discrete variables.

The summary graph captures the independence structure induced by the parent graph for $f_{N|C}$ and contains the added information which of the generating dependencies, indicated by $i \prec\!\!-\!\!-\, k$ in G^V_{par} may be directly or indirectly confounded in the measures of dependence associated with $i \prec\!\!-\!\!-\, k$ in the maximal ancestral graph model for $f_{u|vC}$. For a joint Gaussian density f_V, the induced density $f_{u|vC}$ is again Gaussian, but in general, the form and any parametrization of the density $f_{u|vC}$ induced by f_V may be complex.

Nevertheless, we conjecture that the parameters associated with dashed lines and arrows in G^N_{sum} may often be obtained via a notional stepwise generating process by introducing some additional latent variables that are mutually independent and independent of Y_v, Y_C.

For this, every dashed-line edge $i{-}{-}{-}k$ in the summary graph is replaced by $i \prec\!\!-\!\!-\, \not\!\!\circ \,-\!\!-\!\!\succ k$, each node $\not\!\!\circ$ represents a latent variable and all nodes in $\{\not\!\!\circ\}$ are uncoupled. This generates the summary graph for $f_{u|vC}$ after marginalizing over $\{\not\!\!\circ\}$.

If in a corresponding notional stepwise generating process, the additional latent variables are taken to be discrete and to have a large number of levels, then it should be possible to generate any association corresponding $i{-}{-}{-}k$, or at least approximate it closely. We expect that this follows by using Proposition 5.8 of Studený (2005), or for discrete variables by Theorem 1 of Holland and Rosenberg (1989), but a proof is pending.

4. Generating a summary graph from a larger summary graph. Let a summary graph be given, where the corresponding model, actually or only notionally, arises from a parent graph model by conditioning on Y_c and by marginalising over variables Y_m, $m = (h, k)$, with foster nodes k of c, and outsider nodes h of c.

Then, for a Gaussian triangular system (2.6) in a mean-centered Y with

$$AY = \varepsilon, \qquad \text{cov}(\varepsilon) = \Delta \;\; \text{diagonal}, \qquad A \;\; \text{unit upper-triangular},$$

one obtains with Proposition 2 for $V \setminus \{c, m\} = (\mu, \nu)$ the following equations in $Y_{\mu|c}$, $Y_{\nu|c}$, in the form of equations (3.1),

$$\text{(4.1)} \qquad \begin{pmatrix} B_{\mu\mu} & B_{\mu\nu} \\ 0 & \Sigma^{-1}_{\nu\nu|c} \end{pmatrix} \begin{pmatrix} Y_{\mu|c} \\ Y_{\nu|c} \end{pmatrix} = \begin{pmatrix} \eta'_\mu \\ \zeta_\nu \end{pmatrix}, \quad \text{cov} \begin{pmatrix} \eta'_\mu \\ \zeta_\nu \end{pmatrix} = \begin{pmatrix} W'_{\mu\mu} & 0 \\ . & \Sigma^{-1}_{\nu\nu|c} \end{pmatrix}.$$

The edge matrix components of $G^{V\setminus\{c,m\}}_{\text{sum}}$ are denoted accordingly, by $\mathcal{B}_{\mu\mu}$, $\mathcal{B}_{\mu\nu}$ for arrows pointing to μ, by $\mathcal{W}'_{uu}$ for dashed lines within μ, and by $\mathcal{S}^{\nu\nu.\mu m}$ for full lines within ν.

With added conditioning on a set $c_\nu \subset \nu$, no additional ancestors of c_ν are defined, since every node in ν is already an ancestor of c. But, with added conditioning on $c_\mu \subset \mu$, the set $\mu \setminus c_\mu$ is split into foster nodes f of c_μ and into outsiders o of $r = \{c_\mu, f\}$.

It follows that in the Gaussian model to $G^{V\setminus\{c,m\}}_{\text{sum}}$, equations in Y_o, Y_r given Y_v, Y_c are block-triangular in (o, r). But, by contrast to the split of $V \setminus C$ into (O, R) in equation (3.7), the system is block-triangular but not orthogonal in (o, r) so that conditioning on c_μ in a summary graph is more complex than conditioning in a parent graph.

PROPOSITION 3. Generating $G^{V\setminus\{C,M\}}$ from $G^{V\setminus\{c,m\}}_{\text{sum}}$ and the Gaussian model to $G^{V\setminus\{C,M\}}$ from equations (4.1). *Given equations* (4.1) *to* $G^{V\setminus\{c,m\}}_{\text{sum}}$ *with* $m = (h, k)$ *and new conditioning set* $C = \{c, c_\mu, c_\nu\}$*, and new marginalising set* $M = \{p, q\}$ *with* $p = \{g, h\}$*,* $g \subset o$*, and* $q = \{k, l\}$*,* $l \subset \{f, \nu \setminus c_\nu\}$*, transformed linear parameter matrices and edge matrices are for*

$$r = \mu \setminus o, \qquad \psi = (r, \nu)$$

$$Q = \text{inv}_r\, W'_{\mu\mu},\ \ \mathcal{Q} = \text{zer}_r\, \mathcal{W}'_{\mu\mu},\ \ C_{o\psi} = B_{o\psi} - Q_{or} B_{r\psi},\ \ \mathcal{C}_{o\psi} = \text{In}[\mathcal{B}_{o\psi} + \mathcal{Q}_{or}\mathcal{B}_{r\psi}].$$

For the marginalising set (g, l)*, further transformed linear parameter matrices and edge matrices, arranged in the order* (g, u, l, v)*, be for*

$$u = o \setminus g, \qquad \phi = \psi \setminus \{c_\mu, c_\nu\}, \qquad v = \phi \setminus l,$$

$$K = \text{inv}_{gl} \begin{pmatrix} \tilde{B}_{oo} & \tilde{C}_{o\phi} \\ 0 & \tilde{\Sigma}^{-1}_{\phi\phi|C} \end{pmatrix}, \qquad \mathcal{K} = \text{zer}_{gl} \begin{pmatrix} \tilde{\mathcal{B}}_{oo} & \tilde{\mathcal{C}}_{o\phi} \\ 0 & \tilde{\mathcal{S}}^{\phi\phi.mo} \end{pmatrix}.$$

Then, the Gaussian summary graph model to G^N_{sum}*, which is given by*

$$\begin{pmatrix} K_{uu} & K_{uv} \\ 0 & \Sigma^{-1}_{vv|C} \end{pmatrix} \begin{pmatrix} Y_{u|C} \\ Y_{v|C} \end{pmatrix} = \begin{pmatrix} \eta_u \\ \zeta_{v|q} \end{pmatrix},\ \ \eta_u = (\xi_u - K_{ug}\xi_g) - C_{ul.g}\Sigma_{ll|vC}\zeta_l, \tag{4.2}$$

coincides with the Gaussian model obtained from the triangular system (2.6) *by directly conditioning on* Y_C *and marginalising over* Y_M*.*

The edge matrix components of G^N_{sum}*, which are* $\mathcal{K}_{uu}, \mathcal{K}_{uv}, \mathcal{S}^{vv.Mu}$ *and*

$$\mathcal{W}_{uu} = \text{In}[(\mathcal{Q}_{uu} + \mathcal{K}_{ug}\mathcal{Q}_{gu} + \mathcal{Q}_{ug}\mathcal{K}^{\mathrm{T}}_{ug} + \mathcal{K}_{ug}\mathcal{Q}_{gg}\mathcal{K}^{\mathrm{T}}_{ug}) + \mathcal{C}_{ul.g}\mathcal{S}_{ll|vC}\mathcal{C}^{\mathrm{T}}_{ul.g}], \tag{4.3}$$

coincide with the summary graph obtained from G^V_{par} *by directly conditioning on* C *and marginalising over* M*.*

PROOF. The conditioning set c_μ splits the set of nodes μ into (o, r), where o is without any descendant in $r = \{c_\mu, f\}$ and where every node in f has a descendant in c. This implies a block-triangular form of $B_{\mu\mu}$ in (o, r) in the equations of $Y_{o|\nu c}$ and $Y_{r|\nu c}$, where the residuals η'_o and η'_r are correlated.

For $\psi = (r, \nu)$, block-orthogonality with respect to (o, ψ) in the equations in $Y_{o|c}$ and $Y_{\psi|c}$ is achieved by subtracting from η'_o the value predicted by linear least-squares regression of η'_o on η'_r and ζ_ν. This reduces, because of the orthogonality of the equations in (μ, ν), to subtracting from $Y_{\psi|c}$

$$Q_{or}\eta'_r = Q_{or}B_{r\psi}Y_{\psi|c}.$$

In the resulting equations in $Y_{o|c}$, $Y_{\psi|c}$, the matrix of equation parameters is chosen to be the concentration matrix of $Y_{\psi|c}$ defined by

$$\Sigma^{\psi\psi.mo} = \Sigma^{-1}_{\psi\psi|c} = \begin{pmatrix} B^{\mathrm{T}}_{rr}Q_{rr}B_{rr} & B^{\mathrm{T}}_{rr}Q_{rr}B_{r\nu} \\ . & \Sigma^{-1}_{\nu\nu|c} + B^{\mathrm{T}}_{r\nu}Q_{rr}B_{r\nu} \end{pmatrix}.$$

so that conditioning on a subset of ψ, where $\phi = \psi \setminus \{c_\mu, c_\nu\}$, permits the following transformation without changing independence constraints.

By use of (2.20), the equations in $Y_{o|c}$ are replaced by equations in $Y_{o|C}$, and by use of (2.27), the equations in $Y_{\phi|c}$ by those in $Y_{\phi|C}$. The matrix of equation parameters of $Y_{\phi|C}$ is then $\Sigma^{-1}_{\phi\phi|C}$, the submatrix of $\Sigma^{-1}_{\psi\psi|c}$. The resulting equations give the Gaussian model to the summary graph in node set $V \setminus \{C, m\} = (o, \phi)$. The graph $G^{V\setminus\{C,m\}}$ is defined by the analogously transformed edge matrix components of $G^{V\setminus\{c,m\}}$.

In the Gaussian model to $G^{V\setminus\{C,m\}}$, marginalizing over $Y_{g|C}$, where $g \subset o$, and over $Y_{l|C}$, where $l \subset \phi$, is achieved with partial inversion on g, l. To keep equations in $Y_{u|C}$ and $Y_{v|C}$ no reordering between components of o and ϕ is involved so that block-triangularity is preserved for $u \subset o$ and $v \subset \phi$. Analogously, $G^{V\setminus\{C,M\}}_{\text{sum}}$ is obtained with partial closure on g, l in $G^{V\setminus\{C,m\}}_{\text{sum}}$.

In the resulting equations (4.2), we know by the commutativity (2.12) and exchangeability (2.13) of partial inversion that

$$K_{uu} = [\mathrm{inv}_g \mathrm{inv}_h A]_{u,u} = [\mathrm{inv}_p A]_{u,u},$$

so that $K_{uu} = D_{uu}$, where D is defined for Proposition 2. Furthermore,

$$-K_{uu}\Pi_{u|v.C} = K_{uv} = D_{uv} + D_{uq}\Pi_{q|v.C},$$

so that the equations in $Y_{u|C}$ and $Y_{v|C}$ coincide as given by (4.2) with those given by Proposition 2. The proof is completed by the commutativity and exchangeability property of partial closure, after using the analogous edge matrix expressions and applying the indicator function. □

5. Path interpretations of edge matrix results. The edge matrix results, derived in the previous sections, are now translated into conditions involving specific types of path in summary graphs.

5.1. *Some more terminology and results for graphs.* The inner node in each of the following two-edge paths in summary graphs is called collision node

$$i \longrightarrow \circ \longleftarrow k, \quad i - - - \circ \longleftarrow k, \quad i \longrightarrow \circ - - - k, \quad i - - - \circ - - - k. \tag{5.1}$$

A path is collisionless if it has no inner collision node, it is a pure-collision path if each inner node is a collision node.

Subgraphs induced by three nodes are named V-configurations if they have two edges. The above list contains all possible collision-oriented V-configurations of a summary graph. They share that the inner node is excluded from the conditioning set of any independence constraint on Y_i, Y_k.

In figures of graphs to be modified, we denote conditioning on a node by a boxed in node, $\boxed{\circ}$, and marginalising, as before, by $\not\circ$.

Corollary 1. *The following modifications of the three types of V-configurations in a parent graph generate the three types edge in summary graphs*
(i) i —— k *arises with* $i \longrightarrow \boxed{\circ} \longleftarrow k$,
(ii) $i \longleftarrow k$ *arises with* $i \longleftarrow \not\circ \longleftarrow k$,
(iii) $i - - - k$ *arises with* $i \longleftarrow \not\circ \longrightarrow k$.

Proof. An induced full ik-line is defined in Proposition 2 with $\mathcal{A}_{Ri}^{\mathrm{T}} \mathcal{A}_{Rk}$, an induced arrow with $\mathrm{zer}_i \tilde{\mathcal{A}}$, and an induced dashed line in $\mathcal{A}_{iq} \mathcal{A}_{kq}^{\mathrm{T}}$ in the case that every ancestor in G_{par}^V is also a parent. □

In the relevant family of Gaussian distributions, these induced edges correspond to some nonzero change to a (partial) correlation, see e.g. Wermuth and Cox (2008). Thus, the paths are association-inducing for the family. However, when several paths connect the same variable pair, then for a special member of the given family, the effects of several paths may cancel. Therefore, we speak of edge-inducing instead of association-inducing paths.

For uncoupled nodes $i < k$, the following ik-paths in G_{par}^V could have generated the dashed lines in three of the V-configurations of (5.1); in these ik-paths in G_{par}^V, two arrowheads had met head-on at a collision node, $\circ$,:

$$i \longleftarrow \not\circ \longrightarrow \circ \longleftarrow k, \quad i \longrightarrow \circ \longleftarrow \not\circ \longrightarrow k, \quad i \longleftarrow \not\circ \longrightarrow \circ \longleftarrow \not\circ \longrightarrow k.$$

For all types of V-configurations occurring in a summary graph, the effects are summarized in the following subsections 5.2 and 5.3.

If node k is coupled to node i, then k is named a neighbor of i. A path is said to be chordless if each inner node forms a V-configuration with its two neighbors. Subgraphs induced by n nodes are named U-configurations if they form a chordless path in $n-1$ edges. For four nodes, the following list contains all possible collision-oriented U-configurations of a summary graph

$$i \longrightarrow \circ\, ---\, \circ \longleftarrow k,\quad i\, ---\, \circ\, ---\, \circ \longleftarrow k,\quad i \longrightarrow \circ\, ---\, \circ\, ---\, k,\quad i\, ---\, \circ\, ---\, \circ\, ---\, k.$$

Proposition 4. Orienting a summary graph without foster nodes and double edges. *The independence structure of a summary graph with two types of edge, dashed lines and arrows, and at most one edge for each node pair, cannot be captured by any directed acyclic graph in the same node and edge set if the graph contains a chordless pure-collision path in four nodes.*

Proof. By orienting all undirected edges in such a pure collision path, i.e. by replacing every undirected edge by an arrow, at least one V-configuration results that is no longer collision-oriented. Thereby, a qualitatively different constraint would be introduced for this uncoupled pair. □

Proposition 4 complements a known result for concentration graphs, where a collisionless n-cycle for $n > 3$, i.e. a chordless collisionless path in n nodes having coupled endpoints, cannot be oriented without generating a collision-oriented V-configuration, $\circ \longrightarrow \circ \longleftarrow \circ$, or a directed cycle. The two results together explain why three types of edge may be needed to capture independence structures that result after marginalising and conditioning in G^V_{par}.

5.2. *Edge-inducing paths derived from summary graphs.* The following translation of the edge matrix results of Section 4 shows how additional edges in the summary graph $G^{V\setminus\{C,M\}}_{\text{sum}}$ may be derived from $G^{V\setminus\{c,m\}}_{\text{sum}}$ by checking repeatedly V- and 4-node U-configurations.

For $G^{V\setminus\{c,m\}}_{\text{sum}}$ with ordered node set $N' = V\setminus\{c,m\} = (\mu,\nu)$ of Proposition 3, conditioning on outsider nodes, $c_\mu \subset \mu$, increases the set of foster nodes by splitting $\mu \setminus c_\mu$ into the ordered set (o,f) of remaining outsiders o and additional fosters f and $r = \mu \setminus o = \{c_\mu, f\}$.

Corollary 2. Generating $G^{V\setminus\{C,m\}}_{\text{sum}}$ from $G^{V\setminus\{c,m\}}_{\text{sum}}$. *We let again* $C = \{c, c_\mu, c_\nu\}$, $\psi = (r,\nu)$, *the edge matrix components* $\mathcal{B}_{\mu N'}$, $\mathcal{W}'_{\mu\mu}$, $\mathcal{S}^{\nu\nu.\mu m}$, *and* $\mathcal{Q}_{\mu\mu} = \text{zer}_r \mathcal{W}'_{\mu\mu}$. *One inserts for uncoupled* i,k

1. $i --- k$ *for every* $i --- \boxed{\circ} --- k$ *with* $\boxed{\circ} \in r$; $i,k \in \mu$ $(\mathcal{Q}_{\mu\mu})$;
2. $i \longleftarrow k$ *for every* $i --- \boxed{\circ} \longleftarrow k$ *with* $\boxed{\circ} \in r$; $i \in o$; $k \in \psi$ $(\mathcal{Q}_{or}\mathcal{B}_{r\psi})$;
3. i —— k *for every* $i \longrightarrow \boxed{\circ} --- \boxed{\circ} \longleftarrow k$ *with* $\boxed{\circ} \in r$; $i,k \in \psi$ $(\mathcal{S}^{\psi\psi.mo})$.

Then, one replaces all $i{-}{-}{-}k$ *present between* o *and* r *by* $i \prec\!\!\!-\!\!\!-\, k \in r$, *all edges within* ψ *by full lines and keeps the subgraph induced by nodes* $N' \setminus \{c_\mu, c_\nu\}$. *The graph* $G^{V\setminus\{C,m\}}$ *has node set* $N'' = (o, \phi)$, *with* $\phi = \psi \setminus \{c_\mu, c_\nu\}$.

COROLLARY 3. Generating $G^{V\setminus\{C,M\}}_{\rm sum}$ from $G^{V\setminus\{C,m\}}_{\rm sum}$. *We let again* $M = \{m, g, l\}$ *with* $g \subset o$, $l \subset \phi$, *denote the edge matrix that is block-triangular in* (o, ϕ) *by* $\mathcal{H}''$, $\mathcal{K} = \mathrm{zer}_M \mathcal{H}''$, *and insert for uncoupled nodes* i, k

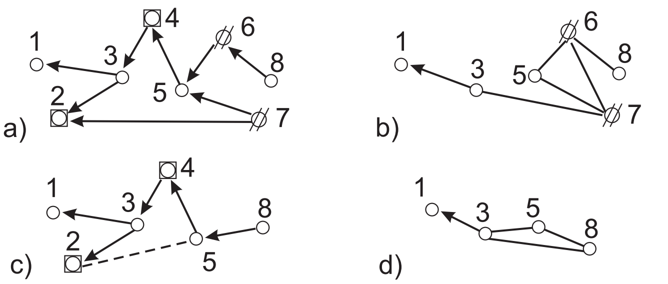
1. $i \,\text{——}\, k$ *for every* $i \,\text{——}\, \not\!\!\not\!\bigcirc \,\text{——}\, k$ *with* $\not\!\!\not\!\bigcirc \in l$; $i, k \subset \phi$ $(\mathrm{zer}_l \mathcal{S}^{\phi\phi.mo})$;
2. $i \prec\!\!\!-\!\!\!-\, k$ *for every* $i \prec\!\!\!-\!\!\!-\, \not\!\!\not\!\bigcirc \prec\!\!\!-\!\!\!-\, k$ *with* $\not\!\!\not\!\bigcirc \in g$; $i, k \in o$; $(\mathrm{zer}_g \mathcal{H}'')$;
   then for every $i \prec\!\!\!-\!\!\!-\, \not\!\!\not\!\bigcirc \,\text{——}\, k$ *with* $\not\!\!\not\!\bigcirc \in l$; $i \in o$; $k \in N''$ $(\mathrm{zer}_{gl} \mathcal{H}'')$.
   Next, one adds for any $i, k \in u$ *for* $\mathcal{W}_{uu}$ *defined in* (4.3)
3. $i{-}{-}{-}k$ *for every* $i \prec\!\!\!-\!\!\!-\, \not\!\!\not\!\bigcirc {-}{-}{-} k$ *and* $i {-}{-}{-} \not\!\!\not\!\bigcirc -\!\!\!-\!\!\!\succ k$ *with* $\not\!\!\not\!\bigcirc \in g$;
   and for every $i \prec\!\!\!-\!\!\!-\, \not\!\!\not\!\bigcirc {-}{-}{-} \not\!\!\not\!\bigcirc -\!\!\!-\!\!\!\succ k$ *with* $\not\!\!\not\!\bigcirc \in g$;
   and for every $i \prec\!\!\!-\!\!\!-\, \not\!\!\not\!\bigcirc \,\text{——}\, \not\!\!\not\!\bigcirc -\!\!\!-\!\!\!\succ k$ *with* $\not\!\!\not\!\bigcirc \in l$.

and keeps the subgraph induced by nodes $N'' \setminus \{g, l\}$.

The following example illustrates stepwise constructions.

EXAMPLE 2. Path constructions of $G^V_{\rm sum}$ for $M = q$ and $p = \emptyset$. *The node set of the parent graph is* $V = (1, \ldots, 8)$. *The conditioning set* $C = \{2, 4\}$, *the marginalising set is* $M = \{6, 7\}$. *The ancestors of* C *outside* C, *i.e. the foster nodes of* C *are in* $F = \{3, 5, 6, 7, 8\}$ *and* $u = O = \{1\}$, $v = \{3, 5, 8\}$.

FIG 4. *a) The generating graph* $G^V_{\rm par}$, *b)* $G^{V\setminus\{C,\emptyset\}}_{\rm sum}$, *c)* $G^{V\setminus\{\emptyset,M\}}_{\rm sum}$, *d)* $G^N_{\rm sum}$.



In this example, the summary graph model defined implicitly with figure 4d) is Markov equivalent to a triangular system of densities in $N = (1, 3, 5, 8)$ even though $G^{V\setminus\{\emptyset,M\}}_{\rm sum}$ in Figure 4c) is not since it contains the chordless pure-collision path $3 -\!\!\!-\!\!\!\succ 2{-}{-}{-}5 \prec\!\!\!-\!\!\!-\, 8$. It is helpful for the planning of small replication studies to know which marginalizing or conditioning leads to standard indpendence straucture s, such as in Figure 4b) and 4d).

5.3. *Some properties of summary graphs.* For any partitioning of the node set $N = V \setminus \{C, M\} = \{\alpha, \beta, \gamma, \delta\}$ of the summary graph, G^N_{sum}, where only γ and δ may be empty sets, the definitions and properties of induced edge matrices imply

$$G^N_{\text{sum}} \Longrightarrow \alpha \perp\!\!\!\perp \beta | C\gamma \iff \mathcal{S}^{\alpha\beta.M\delta} = 0 \iff \mathcal{P}_{\alpha|\beta.C\gamma} = 0 \iff \mathcal{S}_{\alpha\beta|C\gamma} = 0.$$

There are many equivalent path criteria, the one closest to the first criterion for parent graphs, given by Geiger, Verma and Pearl (1990), is the following.

LEMMA 3. Path criterion of global Markov property; Richardson (2003). *A summary graph, G^N_{sum} implies $\alpha \perp\!\!\!\perp \beta | C\gamma$ if and only if it contains no path from node $i \in \alpha$ to node $k \in \beta$ such that of its inner nodes every collision node is in γ or is an ancestor of γ and every other node is outside γ.*

In a summary graph generated by a parent graph, there are two types of path that point to distortions due to indirect confounding in a corresponding maximal ancestral graph model, see Wermuth and Cox (2008). These types of path remain unchanged in a summary graph generated by a multivariate regression chain graph.

With a node named a forefather if it is an ancestor but not a parent and with three dots indicating that there may be more edges of the same type coupling distinct nodes, the following two types of path for for pair (i, k) coupled by an arrow, $i \prec\!\!-\!\!- k$, point to indirect confounding

$$i \,\text{- - -}\, \boxed{\circ} \ldots \boxed{\circ} \,\text{- - -}\, \boxed{\circ} \,\text{- - -}\, k, \qquad i \,\text{- - -}\, \boxed{\circ} \ldots \boxed{\circ} \,\text{- - -}\, \boxed{\circ} \prec\!\!-\!\!- k,$$

where each node $\boxed{\circ}$ along the path is a forefather of offspring node i. In Fig. 2b, the confounding path for $1 \prec\!\!-\!\!- 4$ is $1 \,\text{- - -}\, 3 \prec\!\!-\!\!- 4$.

These warning signals of a summary graph for the induced density $f_{N|C}$ seems to be essential for understanding consequences for $f_{N|C}$ of research hypotheses captured by a parent graph or by a multivariate regression graph.

Acknowledgement. I thank D.R. Cox, G.M. Marchetti and the referees of an earlier version of the paper for their helpful, insightful suggestions.

REFERENCES

BRITO, C. AND PEARL, J. (2002). A new identification condition for recursive models with correlated errors. *Struct. Equ. Model.* **9**, 459–474.

COX, D.R.(2007). *Principles of statistical inference.* Cambridge University Press. Cambridge.

COX, D.R. AND WERMUTH, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Science* **8**, 204–218; 247–277.

Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation*. London: Chapman and Hall.

Cox, D.R. and Wermuth, N. (2004). Causality: a statistical view. *Internat. Statist. Review* **72**, 285-305.

Dawid (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. B* **41**, 1–31.

Dempster, A.P. (1972). Covariance selection. *Biometrics* **28**, 157–175.

Drton, M. (2008). Discrete chain graph models. *Bernoulli*, to appear.

Geiger, D. and Pearl, J. (1993). Logical and algorithmic properties of conditional independence and graphical models *Ann. Statist.* **21**, 2001-2021.

Geiger, D., Verma, T.S. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20**, 507–534.

Holland, Paul W. and Rosenbaum, Paul R. (1989). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14**, 1523–1543.

Koster (2002). Marginalising and conditioning in graphical models. *Bernoulli* **8**, 817–840.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.

Lauritzen, S. L. and Wermuth, N. (1989) Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17**, 31–57.

Marchetti, G.M. and Wermuth, N. (2009). Matrix representations and independencies in directed acyclic graphs. *Ann. Statist.* **46**, to appear.

Richardson, T.S. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Statist.***30**, 145–157.

Richardson, T.S. and Spirtes, P. (2002). Ancestral Markov graphical models. *Ann. Statist.* **30**, 962–1030.

Sadeghi, K. (2008). Graph theoretical proofs of separation criteria. *Thesis, Chalmers University of Technology*

Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica.* **26**, 393–415.

Studený, M. (2005). *Probabilistic Conditional Independence Structures.* Springer Series in Information Science and Statistics, London.

Wermuth, N. and Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli* **4**, 477–495.

Wermuth, N. and Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J.R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 687-717.

Wermuth, N. and Cox, D.R. (2008). Distortion of effects caused by indirect confounding. *Biometrika* **95**, 17–33.

Wermuth, N. and Lauritzen, S.L.(1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discusssion). *J.R. Stat. Soc. Ser. B Stat. Methodol.* **52**, 21-75.

Wermuth, N., Wiedenbeck, M. and Cox, D.R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT, Numerical Mathematics* **46**, 883–901.

Wiedenbeck, M. and Wermuth, N. (2009). Changing parameters by partial mappings. Statistica Sinica, to appear.

Mathematical Statistics
Chalmers/ University of Gothenburg
Chalmers Tvärgata 3
41296 Göteborg, Sweden
E-mail: wermuth@chalmers.se
URL: http://www.math.chalmers.se/∼wermuth