PROBABILITY DISTRIBUTIONS WITH SUMMARY GRAPH STRUCTURE

By Nanny Wermuth *

Mathematical Statistics at Chalmers/ University of Gothenburg, Sweden

A joint density of several variables may satisfy a possibly large set of independence statements, called its independence structure. Often this structure is fully 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 distribution generated. For densities that then result 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 possibly severe distortions of these two types. 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 explain how these lead to special types of path.

1. Introduction. Graphical Markov models are probability distributions defined for a $d_N \times 1$ random vector variable whose components may be discrete or continuous and whose joint density f_V satisfies the 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). With such graphs, each independence constraint is specified for a 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.

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 re-

*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, Recursive regression graph, Triangular system.

peatedly, say over years or decades, and in studies of direct and indirect effects of possible causes on joint responses, see Cox and Wermuth (2004).

For a chain graph, the vertex or node set is $N = \{1, \ldots, d_N\}$ and the edge set consists of pairs i, k of distinct nodes. A partitioning of the node set into so-called chain components, $g(j)$, defines with the ordered sequence $N = (g(1), \ldots, g(j), \ldots, g(J))$ a block-recursive factorization of the joint density, written in a condensed notation in terms of node sets as

$$f_N = \prod_{j=1}^{J} f_{g(j)|g(j+1),\ldots,g(J)}\,. \tag{1.1}$$

Thus, for $N = (a, g(j), b)$ specified in a temporal order of the chain components, $f_{a|g(j)b}$ captures the future and f_b the past of the response node(s) in $g(j)$.

Different types of chain graph result with different types of conditional independence constraints on f_N, separately within and between chain components. Using again only 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. For b the union of two disjoint sets β and γ, written in this paper in a slight abuse of notation as $b = \{\beta, \gamma\} = \beta\gamma$, independence constraints, reflected directly in the factorization of joint density (1.1) and common to all chain graphs, are

$$g(j) \perp\!\!\!\perp \beta|\gamma \iff f_{g(j)|\beta\gamma} = f_{g(j)|\gamma}. \tag{1.2}$$

With multivariate regression chain graphs, a missing edge for node pair $i < k$ constrains the corresponding variable pair in f_N to satisfy

$$i \perp\!\!\!\perp k|b \text{ for } i, k \in g(j), \qquad i \perp\!\!\!\perp k|b \setminus \{k\} \text{ for } i \in g(j), k \in b. \tag{1.3}$$

There are at least two ways to assure that the pairwise independences combine in the same way as in a Gaussian distribution into additional independence constraints to be satisfied by the joint density (1.1). Either this joint density has been generated, possibly via a larger set of variables, as a triangular system, that is via recursive sequence of univariate conditional densities, see Section 2.2. below, or an additional Markov property is assumed, see Kauermann (1996), Richardson (2003), Drton (2008).

In this paper, we treat multivariate regression chain graphs as a subclass of what we name summary graphs. As we shall show, these arise from a given directed acyclic graph by marginalising over a subset M and conditioning on a disjoint subset C of V such that the ordered node set is $N = V \setminus \{M, C\} = (u, v)$ has, loosely speaking, in v the past of C and in u the future of C and a corresponding factorization of the derived density as

$$f_{N|C} = f_{u|vC} f_{v|C}\,. \tag{1.4}$$

Three types of edge, and at most one type of double edge coupling node pairs within u, assure that the summary graph (1) captures each independence implied by the generating graph for $f_{N|C}$ and no other independences and (2) it alerts to possibly severe distortions of generating dependences due to direct or indirect confounding, see Wermuth and Cox (2008).

Though every multivariate chain graph has a generating directed acyclic graph in a larger node set, not every multivariate chain graph model has a generating triangular system with additional, mutually independent variables. The reason is that independences impose typically constraints on permissible parameter values. For instance for a trivariate Gaussian density of the form $f_{1|23}f_2f_3$, the zero correlation $\rho_{23} = 0$ implies a degenerate conditional density $f_{1|23}$ for $\rho_{12} = \rho_{13} = \sqrt{0.5}$ and would lead to a negative conditional variance of Y_1 given Y_2, Y_3 for $\rho_{12}^2 + \rho_{13}^2 > 1$. It is an open problem to characterize the largest subclasses of multivariate regression chain models that can be generated via triangular systems.

With some preliminary results given in Section 2, we define in Section 3 a linear summary graph model and the summary graph. We derive the local Markov properties satisfied by associated probability distributions and generate a summary graph in node set N from a directed acyclic graph in node set $V \supset N$, after conditioning on C and marginalising over nodes M.

In Section 4, we obtain a summary graph from a summary graph given for 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). This method has been used by Marchetti and Wermuth (2008) to prove equivalence of different separation criteria for directed acyclic graphs. It is used in Section 5 to interpret the matrix results in terms of paths.

Directed acyclic graphs after marginalising have been discussed before by Paz (2007) in the form of undirected graphs having additional annotations. Consequences after marginalising and conditioning have also been studied before, by Koster (2002) and by Richardson and Spirtes (2002), who derive, respectively, the classes of MC-graphs and ancestral graphs. In MC-graphs, each node pair may be coupled by up to four types of edge so that corresponding statistical models are frequently under-identified. In ancestral graph models, measures of dependence of the generating process appear frequently in distorted form. In the discussion, Section 6, we relate MC-graphs and ancestral graphs to summary graphs.

2. Notation and preliminary results.

2.1. *Triangular systems of densities and the edge matrix of a parent graph.* For a stepwise process of generating the joint distribution of a vector random variable of dimension $d_V \times 1$ in terms of univariate conditional densities $f_{i|i+1,\dots,d_V}$, 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 $Y_2, \dots, Y_{d_V}$. The conditional densities may be of arbitrary form but are non-degenerate and satisfy precisely the independences captured by the associated graph in d_V nodes (and no others). In this graph, node i represents variable Y_i and each ik-edge a directed relation with Y_i as response to Y_k.

An ik-arrow denotes an edge that starts at node $k > i$ and points to node i. For $i \prec\!\!-\!\!\!- k$, an ik-arrow present, i is named an offspring of k and k a parent of i. The set of parents of node i is denoted by par_i. Each ik-arrow present indicates a non-vanishing conditional dependence of Y_i on Y_k given $Y_{\text{par}_i \setminus k}$. For each node i, the set of missing ik-arrows give

$$f_{i|i+1,\dots,d_V} = f_{i|\text{par}_i} \iff i \perp\!\!\!\perp k|\text{par}_i \text{ for } k > i \text{ not a parent of } i, \tag{2.1}$$

where $f_{i|\text{par}_i} = f_i$ whenever par_i is empty. The joint density factorizes as

$$f_V = \textstyle\prod_{i=1}^{d_V} f_{i|\text{par}_i}. \tag{2.2}$$

The set of independences (2.1) define a directed acyclic graph,which together with the complete ordering of the nodes as $V = (1, \dots, d_V)$, is named the parent graph G^V_{par}. We speak of a joint density generated over G^V_{par} if it has all the features resulting with the just described generating process. For applications, a density generated over a parent graph represents a research hypothesis that is not to be simplified further, since all dependences are included that are large enough to be relevant in a given substantive context; for more discussion see Wermuth and Lauritzen (1990).

Each graph that captures an independence structure has a binary matrix representation with a separate 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

$$\mathcal{A}_{ik} = 1 \text{ if and only if } i \prec\!\!-\!\!\!- k \text{ in } G^V_{\text{par}}. \tag{2.3}$$

Because of the triangular form of the edge matrix $\mathcal{A}$ of G^V_{par}, a density f_V of (2.2), generated over a given parent graph, is also called a triangular system.

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 ik-edge is present in the graph, then node pair ik is said to be coupled; otherwise it is uncoupled.

In an independence graph, every node pair is coupled by at most one edge and each missing ik-edge can be interpreted as an independence constraint on variable pair Y_i, Y_k of the given random vector variable Y, where the conditioning set depends on the type of graph.

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 (2.3), a linear triangular system is a set of recursive linear equations for a mean-centred random vector variable Y of dimension $d_V \times 1$ given by

$$AY = \varepsilon, \tag{2.4}$$

where A is a real-valued $d_V \times d_V$ unit upper-triangular matrix and

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

The vector ε of residuals has zero mean and $\text{cov}(\varepsilon) = \Delta$, a diagonal matrix with $\Delta_{ii} > 0$. The covariance and concentration matrix of Y are, respectively,

$$\Sigma = A^{-1} \Delta A^{-\text{T}}, \qquad \Sigma^{-1} = A^{\text{T}} \Delta^{-1} A.$$

Thus, the linear independences that constrain the equations (2.4) are defined by zeros in the triangular decomposition, (A, Δ^{-1}), of the concentration matrix. The edge matrix of G^V_{par} coincides 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.

The relevant family of Gaussian distributions, i.e. the one generated over G^V_{par}, is defined by (A, Δ^{-1}). This family satisfies no other constraints than the independence structure captured the parent graph, i.e. the constraints given by (2.23) and any additional independence statement that can be derived from this set of independences defining G^V_{par}. All the non-vanishing dependences in this family are specified 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 $G^V_{\rm par}$, probabilistic independence statements combine just like linear independences, see Lemma 1 of Marchetti and Wermuth (2008). There are the following two special properties, the intersection property, i.e. for $V = \{a, b, c, d, e\}$,

$$a \perp\!\!\!\perp b|cd \text{ and } a \perp\!\!\!\perp c|bd \text{ imply } a \perp\!\!\!\perp bc|d$$

and the composition property

$$a \perp\!\!\!\perp b|d \text{ and } a \perp\!\!\!\perp c|d \text{ imply } a \perp\!\!\!\perp bc|d.$$

Therefore, transformations of the edge matrix $\mathcal{A}$, that mimic linear transformations of A, are useful for studying consequences of parent graphs.

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 $V = \{1, \dots, d_V\}$.

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).

Let F and $\mathcal{F}$ be partitioned in the order (a, b). The effect of applying partial closure (2.6) 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

$$\text{inv}_a F \begin{pmatrix} \eta_a \\ Y_b \end{pmatrix} = \begin{pmatrix} Y_a \\ \eta_b \end{pmatrix}. \tag{2.5}$$

DEFINITION 1. Matrix formulations of $\text{inv}_a F$ and $\text{zer}_a \mathcal{F}$.

$$\text{inv}_a F = \begin{pmatrix} F_{aa}^{-1} & -F_{aa}^{-1} F_{ab} \\ F_{ba} F_{aa}^{-1} & F_{bb.a} \end{pmatrix}, \quad \text{zer}_a \mathcal{F} = \text{In}[\begin{pmatrix} \mathcal{F}_{aa}^- & \mathcal{F}_{aa}^- \mathcal{F}_{ab} \\ \mathcal{F}_{ba} \mathcal{F}_{aa}^- & \mathcal{F}_{bb.a} \end{pmatrix}], \tag{2.6}$$

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

and

$$\mathcal{F}_{aa}^- = \text{In}[(n \mathcal{I}_{aa} - \mathcal{F}_{aa})^{-1}], \tag{2.7}$$

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 (2.7) 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}. Thus, $\mathrm{zer}_a\mathcal{F} \geqslant \mathrm{In}[\mathrm{inv}_a F]$ records structural consequences after partial inversion.

It follows directly from (2.5) that F partially inverted on a coincides with F^{-1} partially inverted on $V \setminus a$

$$\mathrm{inv}_a F = \mathrm{inv}_{V\setminus a} F^{-1}\,. \tag{2.8}$$

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

$$\mathrm{inv}_a\,\mathrm{inv}_b F = \mathrm{inv}_b\,\mathrm{inv}_a F, \qquad \mathrm{zer}_a\,\mathrm{zer}_b\,\mathcal{F} = \mathrm{zer}_b\,\mathrm{zer}_a\mathcal{F}, \tag{2.9}$$

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

$$[\mathrm{inv}_a F]_{J,J} = \mathrm{inv}_a F_{JJ}, \qquad [\mathrm{zer}_a\mathcal{F}]_{J,J} = \mathrm{zer}_a\mathcal{F}_{JJ}, \tag{2.10}$$

but partial inversion can be undone while partial closure cannot

$$\mathrm{inv}_{ab}\,\mathrm{inv}_{bc} F = \mathrm{inv}_{ac} F, \qquad \mathrm{zer}_{ab}\,\mathrm{zer}_{\,bc}\,\mathcal{F} = \mathrm{zer}_{abc}\,\mathcal{F}. \tag{2.11}$$

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, with

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

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

$$\mathrm{inv}_a \Sigma^{-1} = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma^{bb.a} \end{pmatrix}, \tag{2.12}$$

where the $\sim$ notation denotes entries in a matrix which is symmetric except for the sign. Property (2.8), $\mathrm{inv}_a\Sigma^{-1} = \mathrm{inv}_b\Sigma$, leads at once to several well *known dual expressions for the three submatrices in* (2.12)*, 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 equations derived from a linear system in a mean-centred vector variable Y with covariance matrix Σ, that is block-triangular in two ordered blocks (a, b), so that

$$HY = \eta, \text{ with } H_{ba} = 0, \quad \text{cov}(\eta) = W \text{ positive definite,} \tag{2.13}$$

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

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

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

$$\text{inv}_a(H^{\mathrm{T}}W^{-1}H) = \begin{pmatrix} K_{aa}Q_{aa}K_{aa}^{\mathrm{T}} & K_{ab} + K_{aa}Q_{ab}K_{bb} \\ \sim & H_{bb}^{\mathrm{T}}Q_{bb}H_{bb} \end{pmatrix} = \begin{pmatrix} \Sigma_{aa|b} & \Pi_{a|b} \\ \sim & \Sigma_{bb}^{-1} \end{pmatrix}. \tag{2.14}$$

Identification is an issue for the block-triangular equations (2.13), but it is assured under the general condition that for any pair ik within a either $H_{ik} = 0$ or $W_{ik} = 0$, see e.g. Brito and Pearl (2002). In particular, within the subclasses of linear multivariate regression chains and of linear triangular systems, each model is identified, i.e. all its parameters are estimable.

This key result for summary graphs also specializes to the following.

COROLLARY 1. Correlated equations induced by reordering in linear triangular systems (2.4). *For the linear triangular systems* (2.4)*, let set a be any subset of V, $b = V \setminus a$, and the order within both, a and b, preserved from V. From $AY = \varepsilon$, the matrix $\tilde{A}$ is A partitioned according to (a, b) and we let $B = \text{inv}_a\tilde{A}$, $\eta_b = \varepsilon_b - B_{ba}\varepsilon_a$ and*

$$W = \text{cov}\begin{pmatrix} \varepsilon_a \\ \eta_b \end{pmatrix} = \begin{pmatrix} \Delta_{aa} & -\Delta_{aa}B_{ba}^{\mathrm{T}} \\ . & \Delta_{bb} + B_{ba}\Delta_{aa}B_{ba}^{\mathrm{T}} \end{pmatrix}, \quad Q = \text{inv}_b W.$$

Then, the induced submatrices of $\text{inv}_a\tilde{\Sigma}^{-1}$ *are*

$$\Sigma_{aa|b} = B_{aa}Q_{aa}B_{aa}^{\mathrm{T}}, \quad \Pi_{a|b} = B_{ab} + B_{aa}Q_{ab}B_{bb}, \quad \Sigma^{bb.a} = B_{bb}^{\mathrm{T}}Q_{bb}B_{bb}. \tag{2.15}$$

PROOF. After partial inversion of $\tilde{A}$ on a, the linear triangular system is modified to

$$B\begin{pmatrix}\varepsilon_a\\ Y_b\end{pmatrix} = \begin{pmatrix}Y_a\\ \varepsilon_b\end{pmatrix}.$$

Associated block-triangular equations in (a, b), that leave the equation in Y_a unchanged but introduce correlations among ε_a and η_b, are

$$A_{aa}Y_a + \tilde{A}_{ab}Y_b = \varepsilon_a, \qquad B_{bb}Y_b = \eta_b,$$

so that Lemma 1 applies with Q as defined above and K the upper block-triangular part of B with respect to (a, b). □

In Corollary 1, the modified residuals $\eta_b = \varepsilon_b - B_{ba}\varepsilon_a$ remove the influence of Y_a on the equations in Y_b. At the same time, correlated residuals are, among others, introduced in the equation of Y_a, since $\text{cov}(\varepsilon_a, \eta_b) = -\Delta_{aa}B_{ba}^{\mathrm{T}}$ so that the residuals ε_a are correlated with Y_b, the explanatory variable in the equation of response variable Y_a.

In econometrics, such a response variable is called endogenous. One speaks also of an endogeneity problem, since distorted least squares estimates of dependences result if such strong correlations are not recognized, see also equation (3.4) below and Wermuth and Cox (2008).

COROLLARY 2. Orthogonal equations induced by reordering and conditioning in linear triangular systems (2.4); Corollary 1 of Wermuth and Cox (2004). *The three induced matrices in equation* (2.15) *coincide with those implied by the following triangular equations in* $Y_{a|b}$ *and* Y_b *with residuals* η_a *uncorrelated of* η_b

$$A_{aa}Y_{a|b} = \eta_a, \qquad B_{bb}Y_b = \eta_b, \qquad \textit{with} \quad \eta_b = \varepsilon_b - B_{ba}\varepsilon_a, \qquad \eta_a = \varepsilon_a - Q_{ab}\eta_b.$$

PROOF. The residuals $\eta_a = \varepsilon_a - Q_{ab}\eta_b$ are uncorrelated with η_b, since Q_{ab} is the matrix of coefficients of η_b in linear least-squares regression of ε_a on η_b. The random variable $Y_{a|b}$ is defined in terms of $\Pi_{a|b}$ of equation (2.15) and has residual covariance $\text{cov}(\eta_a) = Q_{aa}$. □

Thus, from both the correlated equations in Y_a and Y_b leading to Corollary 1 and the orthogonal equations in $Y_{a|b}$ and Y_b of Corollary 2, the concentration matrix of Y may be recovered with $\text{inv}_a(\text{inv}_a\Sigma^{-1})$ and hence also the generating triangular system via the triangular decomposition (A, Δ^{-1}) of Σ^{-1}. Therefore, the transformations from the parameters in equations (2.4) to the parameters in the equations of Corollaries 1 and 2 are both one-to-one. One important consequence is that the independence structure of the relevant family of Gaussian densities remains unchanged by both transformations.

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

LEMMA 2. The edge matrix components induced for $\text{inv}_a\Sigma^{-1}$ by the edge matrices $\mathcal{H}, \mathcal{W}$ of the block-triangular system (2.13). *Structural zeros of* $\text{inv}_a\Sigma^{-1}$ *are given by zeros in*

$$\begin{pmatrix} \mathcal{S}_{aa|b} & \mathcal{P}_{a|b} \\ . & \mathcal{S}^{bb.a} \end{pmatrix} = \text{In}[\begin{pmatrix} \mathcal{K}_{aa}Q_{aa}\mathcal{K}_{aa}^{\text{T}} & \mathcal{K}_{ab} + \mathcal{K}_{aa}Q_{ab}\mathcal{K}_{bb} \\ . & \mathcal{H}_{bb}^{\text{T}}Q_{bb}\mathcal{H}_{bb} \end{pmatrix}], \tag{2.16}$$

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 linear parameter submatrices of $\text{inv}_a\Sigma^{-1} = H^{\text{T}}W^{-1}H$ in Lemma 1 are expressed without any self-cancelling matrix operations such as a matrix multiplied by its inverse. When such linear parameter submatrices are replaced by non-negative submatrices of $\mathcal{H}, \mathcal{W}$, having the appropriate structural zeros, then just the structural zeros in $\text{inv}_a\Sigma^{-1}$ are preserved by multiplying, summing and applying the indicator function, see also Lemma 3 of Marchetti and Wermuth (2008). □

By contrast to the induced edge matrix components in (2.16), $\text{inv}_a\Sigma^{-1}$ of (2.14) may contain additional zeros due to special parametric constellations.

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 $i \perp\!\!\!\perp k|b \setminus k$.

One direct application of Lemma 2 is to the equations of Corollary 1.

COROLLARY 3. Equations (5.2), Wermuth, Wiedenbeck and Cox (2006). Edge matrices induced by a parent graph after reordering and conditioning. *For a given parent graph with edge matrix $\mathcal{A}$ in equation* (2.3)*, for a nonempty subset a of V, $b = V \setminus a$, and the order within both a and b preserved from V, we denote by $\tilde{\mathcal{A}}$ the matrix $\mathcal{A}$ partitioned according to (a, b), and let*

$$\mathcal{B} = \mathrm{zer}_a \tilde{\mathcal{A}}, \qquad \mathcal{W} = \mathrm{In}[\begin{pmatrix} \mathcal{I}_{aa} & \mathcal{B}_{ba}^{\mathrm{T}} \\ . & \mathcal{I}_{bb} + \mathcal{B}_{ba}\mathcal{B}_{ba}^{\mathrm{T}} \end{pmatrix}].$$

The edge matrix components induced by G^V_{par} for $\mathcal{S}_{aa|b}$, $\mathcal{P}_{a|b}$, and $\mathcal{S}^{bb.a}$ result with equations (2.16) *and $\mathcal{H}$ the upper block-triangular part of $\mathcal{B}$ with respect to (a, b), $\mathcal{Q} = \mathrm{zer}_b \mathcal{W}$.*

Corollary 3 implies for $i \in a$ and $k \in b$, that no ik-edge present in the induced covariance graph with edge matrix $\mathcal{W}_{aa}$ coincides with an ik-edge present in the induced directed graph with edge matrix $\mathcal{H}_{aa}$, since

$$\mathcal{B}_{ab}^{\mathrm{T}} = 0 \iff \mathcal{W}_{ab} = 0 \iff \mathcal{Q}_{ab} = 0, \qquad \mathcal{B}_{ab} = 0 \iff \mathcal{H}_{ab} = 0.$$

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

$$Y_a = \Pi_{a|b} Y_b + \epsilon_a, \ \text{with } \mathrm{E}(\epsilon_a) = 0, \ \mathrm{cov}(\epsilon_a, Y_b) = 0. \tag{2.17}$$

For residuals ϵ_a uncorrelated with the regressor variables Y_b, taking expectations in the equation given by $Y_a Y_b^{\mathrm{T}}$ defines $\Pi_{a|b}$ via $\Sigma_{ab} = \Pi_{a|b}\Sigma_{bb}$ and $\Sigma_{aa|b} = \mathrm{cov}(\epsilon)$, as before in Example 1.

We write the 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 (2.17) 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

$$Y_{a|\gamma} = \Pi_{a|\beta.\gamma} Y_{\beta|\gamma} + \epsilon_a, \ \text{with } \mathrm{E}(\epsilon_a) = 0, \ \mathrm{cov}(\epsilon_a, Y_{\beta|\gamma}) = 0, \tag{2.18}$$

with residuals ϵ_a unchanged compared to model (2.17).

This may be proven using Example 1. By moving from $\text{inv}_\gamma\Sigma$ to $\text{inv}_b\Sigma = \text{inv}_\beta(\text{inv}_\gamma\Sigma)$, 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 for $Y_{a|b}$ with $\Pi_{a|\beta.\gamma} = \Sigma_{a\beta|\gamma}\Sigma^{-1}_{\beta\beta|\gamma}$ and

$$\Pi_{a|\gamma.\beta} = \Pi_{a|\gamma} - \Pi_{a|\beta.\gamma}\Pi_{\beta|\gamma}. \tag{2.19}$$

Equation (2.19) 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 (2.19) implies, for all distributions generated over a parent graph

$$\mathcal{P}_{a|\gamma.\beta} = \mathcal{P}_{a|\gamma} \text{ if } \mathcal{P}_{a|\beta.\gamma} = 0. \tag{2.20}$$

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 and orthogonal in $N = (a, b)$. Then, we have in equations (2.13) $\text{cov}(\eta_a, \eta_b) = 0$ added to $H_{ba} = 0$ so that $0 = W_{ab} = W^{\mathrm{T}}_{ba}$.

After partial inversion on a, the linear multivariate regression model (2.17) with

$$\Pi_{a|b} = -H^{-1}_{aa}(H_{a\beta} \quad H_{a\gamma}), \tag{2.21}$$

results, see equations (2.14). In econometrics, such models have been named reduced form equations. An analogue of Cochran's equation (2.19) is then

$$-H_{aa}\Pi_{a|\gamma} = H_{a\gamma} + H_{a\beta}\Pi_{\beta|\gamma}. \tag{2.22}$$

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

$$\alpha \perp\!\!\!\perp \beta|\gamma \iff \mathcal{P}_{\alpha|\beta.\gamma} = 0 \iff \mathcal{H}_{\alpha\beta.\delta} = 0. \tag{2.23}$$

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

$$\Sigma^{-1}Y = \zeta \text{ with } \text{cov}(Y) = \Sigma, \tag{2.24}$$

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

$$Y_a = \Pi_{a|b} Y_b + \Sigma_{aa|b}\zeta_a, \qquad \Sigma^{bb.a} Y_b = \zeta_{b|a}, \qquad \zeta_{b|a} = \zeta_b + \Pi_{a|b}^{\mathrm{T}}\zeta_a, \tag{2.25}$$

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

$$\alpha \perp\!\!\!\perp \beta | \gamma \iff \mathcal{S}^{\alpha\beta.\delta} = 0. \tag{2.26}$$

2.7. *Markov properties of parent graphs.* Equivalent ways of characterizing the independence structure given by a graph, in terms of so-called pairwise, local and global statements, have for instance been studied by Pearl and Paz (1987) for undirected full-line graphs and by Lauritzen et al. (1990) for directed acyclic graphs, see also Lauritzen (1996). For parent graphs, there are in our notation the following three equivalent characterizations.

LEMMA 3. Equivalent Markov properties of parent graphs. *For a parent graph in node set $V = (1, \ldots, d) = \{a, b\}$ with edge matrix $\mathcal{A}$, let $\emptyset \neq \alpha = a \setminus \delta$, $\emptyset \neq \beta = b \setminus \gamma$ and $\mathcal{B} = \mathrm{zer}_a \tilde{\mathcal{A}}$. Then, the following Markov properties are equivalent*

(*i*) *pairwise: for each $i \in V$ and $k > i$ not a parent of i: $i \perp\!\!\!\perp k | \{i+1, \ldots, d\} \setminus \{k\} \iff \mathcal{A}_{ik} = 0$,*
(*ii*) *local: for each $i \in V$ and $k > i$ not a parent of i: $i \perp\!\!\!\perp k | \mathrm{par}_i \iff \mathcal{A}_{i,\{i+1,\ldots,d\}\setminus \mathrm{par}_i} = 0$,*
(*iii*) *global: $\alpha \perp\!\!\!\perp \beta | \gamma \iff \mathrm{In}[\mathcal{B}_{\alpha\beta} + \mathcal{B}_{\alpha a}\mathcal{B}_{ba}^{\mathrm{T}}(\mathcal{I}_{bb} + \mathcal{B}_{ba}\mathcal{B}_{ba}^{\mathrm{T}})^{-}\mathcal{B}_{b\beta}] = 0$.*

PROOF. The edge matrix $\mathcal{P}_{\alpha|\beta.\gamma}$ of Corollary 3 is expressed in (*iii*) with submatrices of $\mathcal{B}$. Since $\mathcal{B}$ is obtained in terms of $\mathcal{A}$, defined in (2.3) and equivalently in (*ii*), (*ii*) $\implies$ (*iii*). For $\alpha = \{i\}$, an order-compatible split for (a, b) is defined with equation (*i*) of Lemma 3 for $\beta = \{k\}$ and $\gamma = \{i+1, \ldots, d\} \setminus \{k\}$. Equation (2.23) applies to (i, γ) and (*iii*) $\implies$ (*i*). Also equation (2.20) applies. Using it in sequence, on nodes $k > i$ and $k \notin \mathrm{par}_i$, reduces the conditioning set in (*i*) to the conditioning set in (*ii*), so that (*i*) $\implies$ (*ii*). ☐

3. Linear summary graph models and summary graphs. As we shall see, summary graphs have in contrast to parent graphs no general pairwise Markov property, but local and global Markov properties and linear summary graph models are special block-triangular equations (2.13).

3.1 Definitions and Markov properties. As is to be described in more detail in the next section, starting from a linear triangular system (2.4) 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 in the future of Y_C and equations in $Y_{N|C}$ for $N = (u, v)$ of the following form.

DEFINITION 2. Linear summary graph model. *A linear summary graph model, that is generated by marginalising over Y_M and conditioning on Y_C in the linear triangular system* (2.4)*, is the following system of equations $HY_{N|C} = \eta$, that is block-triangular and orthogonal in* (u, v)*,*

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

where H_{uu} is unit upper-triangular. Additional zero constraints on H and $\mathrm{cov}(\eta)$ *are captured by the graph denoted by G^N_{sum}, defined below.*

Equation (3.1) implies for $Y_{u|C}$ given $Y_{v|C}$ a constrained multivariate regression model (2.17) and for $Y_{v|C}$ a linear concentration graph model, which had been introduced for joint Gaussian distributions under the name of covariance selection by Dempster (1972). The residuals of $Y_{u|C}$ and $Y_{v|C}$ are uncorrelated so that by (2.21), $\Pi_{u|v.C} = -H^{-1}_{uu} H_{uv}$ and by (2.18), the equations in $Y_{u|C}$ of (3.1) can also be written as

$$(3.2)\qquad H_{uu} Y_{u|vC} = \eta_u \text{ with } W_{uu} = H_{uu} \Sigma_{uu|vC} H^{\mathrm{T}}_{uu}.$$

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 a triangular system (2.4) of unit upper-triangular form, but by contrast to a triangular system, the residuals η_u are in general correlated.

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 3. 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*,*

(3.3) $\qquad \mathcal{W}_{uu}$, *symmetric, for dashed lines within* u,
$\qquad \mathcal{S}^{vv.uM}$, *symmetric, for full lines within* v.

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 the linear 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* $\Sigma^{vv.uM} = \Sigma^{-1}_{vv|C}$ *of the linear summary graph model* (3.1), *respectively.*

For density $f_{N|C}$ of $Y_{N|C}$, obtained from the relevant family of Gaussian distribution of Y_V, the cardinality of the set of edges missing in $G^{V\setminus\{C,M\}}_{\text{sum}}$ gives the number of parameters in equations (3.1) implied to be zero by G^V_{par}.

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 $Y_{d|vC}$, where d is the set of nodes in u larger than i. This gives via least-squares regression of residual η_i on η_d

$$(3.4) \quad -\Pi_{i|d.Cv} = H_{id} - W_{id}W^{-1}_{dd}H_{dd} \text{ and } \mathcal{P}_{i|d.Cv} = \text{In}[\mathcal{H}_{id} + \mathcal{W}_{id}\mathcal{W}^{-}_{dd}\mathcal{H}_{dd}].$$

One consequence is that for an uncoupled node pair ik with $k \in d$, no independence statement 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}.

PROPOSITION 1. Local Markov properties of summary graphs. *Let* β *denote subsets of* N *uncoupled to node* i *in* G^N_{sum} *of* (3.3) *which has edge matrix components* $\mathcal{H}_{uN}$, $\mathcal{W}_{uu}$ *and* $\mathcal{S}^{vv.uM}$. *Let further* d *contain all nodes larger than* i *in* $\mathcal{H}_{uu}$, *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 (3.4), (ii) with (2.23) and (iii) with (2.26). □

Characterisation of different types of path in a summary graph, given below in Section 5, lead to interpretations of these and later matrix conditions.

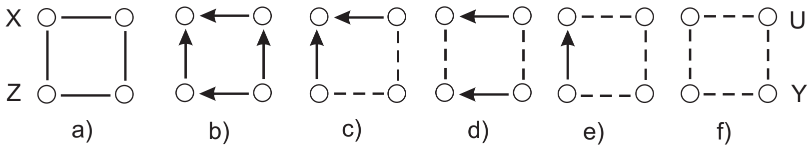
COROLLARY 4. Order-compatible consequences of summary graphs. *A summary graph* G^N_{sum} *of* (3.3), *with edge matrix components* $\mathcal{H}_{uN}, \mathcal{W}_{uu}, \mathcal{S}^{vv.Mu}$, *and* $\mathcal{K}_{uu} = \text{zer}_u\mathcal{H}_{uu}$, *implies*

(i) *for* $\alpha, \beta \in v : \alpha \perp\!\!\!\perp \beta | Cv \setminus \{\alpha, \beta\} \iff \mathcal{S}^{\alpha\beta.Mu} = 0$,
(ii) *for* $\alpha \in u$ *and* $\beta \in v : \alpha \perp\!\!\!\perp \beta | Cv \setminus \beta \iff \mathcal{H}_{\alpha\beta.u\setminus\alpha} = 0$,
(iii) *for* $\alpha, \beta \in u : \alpha \perp\!\!\!\perp \beta | Cv \iff \mathcal{K}_{\alpha u} \mathcal{W}_{uu} \mathcal{K}_{\beta u}^{\mathrm{T}} = 0$.

PROOF. The results follow with the local Markov property (iii) in Proposition 1 for (i), equation (2.23) for (ii), and Lemma 2 for (iii) by using $\mathcal{W}_{uv} = 0$. □

The following figure shows special cases of summary graphs that result after 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.

FIG 1. *Consequences of Corollary 4. 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).*



Substantive research examples with well-fitting data to linear models of Figure 1 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).

3.2 Generating G^N_{sum} *from* G^V_{par} *by first conditioning.* Starting from a linear triangular system in (2.4) 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 (3.5). 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. The following equations are orthogonal and block-triangular in (O, R)

$$\begin{pmatrix} A_{OO} & A_{OR} \\ 0 & A_{RR} \end{pmatrix} \begin{pmatrix} Y_O \\ Y_R \end{pmatrix} = \begin{pmatrix} \varepsilon_O \\ \varepsilon_R \end{pmatrix}. \tag{3.5}$$

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 linear triangular system* (2.4) *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*

$$H_{uu} = D_{uu}, \quad H_{uv} = D_{uv} + D_{uq}\Pi_{q|v.C}, \tag{3.6}$$

$$W_{uu} = (\Delta_{uu} + D_{up}\Delta_{pp}D^{\mathrm{T}}_{up}) + (D_{uq}\Sigma_{qq|vC}D^{\mathrm{T}}_{uq}); \tag{3.7}$$

the induced edge matrix components of the summary graph G^N_{sum} *are*

$$\mathcal{H}_{uu} = \mathcal{D}_{uu}, \quad \mathcal{H}_{uv} = \text{In}[\mathcal{D}_{uv} + \mathcal{D}_{uq}\mathcal{P}_{q|v.C}], \tag{3.8}$$

$$\mathcal{W}_{uu} = \text{In}[(\mathcal{I}_{uu} + \mathcal{D}_{up}\mathcal{D}^{\mathrm{T}}_{up}) + (\mathcal{D}_{uq}\mathcal{S}_{qq|vC}\mathcal{D}^{\mathrm{T}}_{uq})]. \tag{3.9}$$

Proof. From the equations (3.5) in Y, the equations in $Y_{O|C}$ and $Y_{F|C}$

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

are obtained by using (2.18), (2.24) and (2.25). Then, equations (3.1) in $Y_{u|C}$, $Y_{v|C}$ result, with parameters given in (3.6), (3.7), 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

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

After replacing the defining matrix components in (3.6), (3.7) by their corresponding edge matrix components and applying the indicator function, the induced edge matrix components (3.8), (3.9) of G^N_{sum} are obtained. □

It is instructive to also check the relations of the parameter matrices in (3.6), (3.7) to multivariate regression coefficients and to conditional covariance matrices. With $\Pi_{u|R} = -D_{uu}^{-1}(D_{uv},\ D_{uq},\ D_{uC})$, equation (2.22) 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 (3.7)

$$D_{uu}^{-1}W_{uu}D_{uu}^{-T} = \Sigma_{uu|vqC} + \Pi_{u|q.vC}\Sigma_{qq|vC}\Pi_{u|q.vC}^{\mathrm{T}} = \Sigma_{uu|vC},$$

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

3.3 Generating G^N_{sum} from G^V_{par} by first marginalising. In Proposition 2, conditioning on C was followed by marginalising over M. By marginalising instead first over M and removing M and then conditioning on C and removing C, the induced matrices in Propositions 2 can be expressed in a different but equivalent way.

COROLLARY 5. Equations and graph obtained after marginalising over Y_M, then conditioning on Y_C. *Let the node sets and matrices D, $\mathcal{D}$ and W_{uu}, $\mathcal{W}_{uu}$ be as in Proposition 2, and the sets $w = \{v, C\}$, $R = \{w, q\}$. Additional analogously transformed parameter and edge matrices be*

$$E_{RR} = \mathrm{inv}_q\tilde{A}_{RR}, \quad \mathcal{E}_{RR} = \mathrm{zer}_q\tilde{\mathcal{A}}_{RR}, \quad Q_{RR} = \mathrm{inv}_w\tilde{W}_{RR}, \quad \mathcal{Q}_{RR} = \mathrm{zer}_w\tilde{\mathcal{W}}_{RR},$$

where W_{RR} is the joint covariance matrix of ε_q and $\varepsilon_w - E_{wq}\varepsilon_q$, so that

$$W_{RR} = \begin{pmatrix} \Delta_{qq} & -\Delta_{qq}E_{wq}^{\mathrm{T}} \\ \sim & \Delta_{ww} + E_{wq}\Delta_{qq}E_{wq}^{\mathrm{T}} \end{pmatrix}, \qquad \mathcal{W}_{RR} = \mathrm{In}[\begin{pmatrix} \mathcal{I}_{qq} & \mathcal{E}_{wq}^{\mathrm{T}} \\ . & \mathcal{I}_{ww} + \mathcal{E}_{wq}\mathcal{E}_{wq}^{\mathrm{T}} \end{pmatrix}].$$

Then for $F \subset R$, the submatrices of $\mathrm{inv}_q(A_{RR}^{\mathrm{T}}W_{RR}^{-1}A_{RR})$, $\mathrm{zer}_q(\mathcal{A}_{RR}^{\mathrm{T}}\mathcal{W}_{RR}^{-}\mathcal{A}_{RR})$ *are, respectively,*

$$\begin{pmatrix} \Sigma_{qq|vC} & \Pi_{q|v.C} \\ \sim & \Sigma^{vv.qO} \end{pmatrix}, \qquad \begin{pmatrix} \mathcal{S}_{qq|vC} & \mathcal{P}_{q|v.C} \\ . & \mathcal{S}^{vv.qO} \end{pmatrix}.$$

Used in (3.6) (3.7) *and in* (3.8), (3.9), *one obtains the parameter matrices in the linear summary graph model and the edge matrix components to G^N_{sum}.*

PROOF. Partial inversion on p, q gives $D_{uw} + D_{uq}E_{qw}$ as equation parameter of Y_w, and $W_{uw} = D_{uq}E_{qq}\Delta_{qq}E_{wq}^{\mathrm{T}}$ as covariance among the residuals of Y_u and Y_w. This covariance is removed in an orthogonalising step.With Lemma 1, $\mathrm{inv}_q(A_{RR}^{\mathrm{T}}W_{RR}^{-1}A_{RR})$ is expressed in terms of E and Q. With Lemma 2, the corresponding induced edge matrix components use $\mathcal{E}$ and $\mathcal{Q}$. □

4. Generating a summary graph from a summary graph. Let a linear summary graph or a multivariate regression 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, the starting linear parent graph model is the triangular system of equation (2.4) in a mean-centred variable Y where

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

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

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

and the edge matrix components of $G^{V\setminus\{c,m\}}_{\text{sum}}$, 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. For the linear equations (4.1), these represent the induced edge matrices of $B_{\mu\mu}, B_{\mu\nu}, W'_{\mu\mu}$, and of $\Sigma^{\nu\nu.\mu m} = \Sigma^{-1}_{\nu\nu|c}$, respectively.

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 linear 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 (3.5), the system is block-triangular but not orthogonal in (o, r) so that conditioning on c_μ in a summary graph is different from conditioning in a parent graph.

PROPOSITION 3. Generating $G^{V\setminus\{C,M\}}$ from $G^{V\setminus\{c,m\}}_{\text{sum}}$ and the linear 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* $q = \{k, l\}$, $l \subset \{f, \nu \setminus c_\nu\}$, *let transformed linear parameter matrices and edge matrices be for*

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

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

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

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

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

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

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

coincides with the linear model obtained from the triangular system (2.4) *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

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

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 equations 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 (2.18), the equations in $Y_{o|c}$ are replaced by equations in $Y_{o|C}$, and by use of (2.25), 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 are the linear 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 linear 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, $G_{\text{sum}}^{V\setminus\{C,M\}}$ is obtained with partial closure on g, l in $G_{\text{sum}}^{V\setminus\{C,m\}}$.

In the resulting equations (4.2), we know by the commutativity (2.9) and exchangeability (2.10) of partial inversion that

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

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 named collision node

(5.1) $$i \longrightarrow\!\!\!\!\succ \circ \prec\!\!\!\!\longleftarrow k, \quad i - - - \circ \prec\!\!\!\!\longleftarrow k, \quad i \longrightarrow\!\!\!\!\succ \circ - - - k, \quad i - - - \circ - - - k.$$

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 by a crossed out node, $\not\circ$.

Corollary 6. *The following modifications of the three types of V-configurations in a parent graph generate the three types edge in summary graphs*
(i) $i \,\text{---}\, k$ *arises with* $i \longrightarrow\!\!\!\!\succ \boxed{\circ} \prec\!\!\!\!\longleftarrow k$,
(ii) $i \prec\!\!\!\!\longleftarrow k$ *arises with* $i \prec\!\!\!\!\longleftarrow \not\circ \prec\!\!\!\!\longleftarrow k$,
(iii) $i - - - k$ *arises with* $i \prec\!\!\!\!\longleftarrow \not\circ \longrightarrow\!\!\!\!\succ k$.

PROOF. An induced full ik-line is defined in Proposition 2 with $\mathcal{A}_{Ri}^{\mathrm{T}}\mathcal{A}_{Rk}$, an additional full arrow with $\mathrm{zer}_i\tilde{\mathcal{A}}$, and an induced dashed line in Corollary 5 with $\mathcal{E}_{iq}\mathcal{E}_{kq}^{\mathrm{T}}$ which reduces to $\mathcal{A}_{iq}\mathcal{A}_{kq}^{\mathrm{T}}$ if every ancestor in G_{par}^{V} is also a parent. □

In the relevant family of Gaussian distributions, these induced edges correspond to some nonzero amount added to the (partial) correlation defined by the graph for Y_i, Y_k, see e.g. Wermuth and Cox (2008). However, whenever there are several paths connecting the same variable pair, for a given member of the family the effects may cancel, see Wermuth and Cox (1998). Therefore, one speaks here in general only of edge-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, ∘,:

$$i \prec\!\!\!-\!\!\!-\, \oslash \,-\!\!\!-\!\!\!\succ \circ \prec\!\!\!-\!\!\!- k, \quad i \,-\!\!\!-\!\!\!\succ \circ \prec\!\!\!-\!\!\!-\, \oslash \,-\!\!\!-\!\!\!\succ k, \quad i \prec\!\!\!-\!\!\!-\, \oslash \,-\!\!\!-\!\!\!\succ \circ \prec\!\!\!-\!\!\!-\, \oslash \,-\!\!\!-\!\!\!\succ k.$$

For all types of V-configurations occuring in a summary graph, the effects are summarized in the following subsections.

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 ⊔-configurations if they form a chordless path in $n-1$ edges. For four nodes, the following list contains all possible collision-oriented ⊔-configurations of a summary graph

$$i \,-\!\!\!-\!\!\!\succ \circ - - - \circ \prec\!\!\!-\!\!\!- k, \quad i - - - \circ - - - \circ \prec\!\!\!-\!\!\!- k, \quad i \,-\!\!\!-\!\!\!\succ \circ - - - \circ - - - k, \quad i - - - \circ - - - \circ - - - k.$$

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, $\circ \,-\!\!\!-\!\!\!\succ \circ \prec\!\!\!-\!\!\!- \circ$, 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 parent graphs.* The edge matrix results of Section 3 are now shown to reduce to local operations on paths. For this, we exploit that the transpose of an edge matrix for arrows indicates the change of direction of the arrows, see the proof of Corollary 6. As a convention we use that $\text{zer}_a\mathcal{F}$ preserves the type of edge of $\mathcal{F}$ whenever $\mathcal{F}$ represents only one type of edge, such as arrows in $\tilde{\mathcal{A}}$, full line lines in $\tilde{\mathcal{S}}^{FF.O}$ or dashed lines in $\tilde{\mathcal{W}}_{uu}$. In the following Corollaries, the justifying matrix results of Section 3 are listed in brackets.

COROLLARY 7. Generating $G^{V\setminus\{C,\emptyset\}}_{\text{sum}}$ from G^V_{par}. *We let again* $V = (O, R)$, $R = \{C, F\}$, *and* $\mathcal{A}$ *the edge matrix of* G^V_{par}. *One inserts for uncoupled* i, k

$i \frac{\quad}{} k$ *for every* $i \longrightarrow \boxcircle \longleftarrow k$ *with* $\boxcircle \in R$; $i, k \in F$ ($\mathcal{A}^{\mathrm{T}}_{RR}\mathcal{A}_{RR}$),

then replaces edges in F *by full lines and removes all edges and nodes of* C.

COROLLARY 8. Generating $G^{V\setminus\{C,M\}}_{\text{sum}}$ from $G^{V\setminus\{C,\emptyset\}}_{\text{sum}}$. *We let again* $M = (p, q)$, $u = O \setminus p$, $v = F \setminus q$, $\mathcal{D} = \text{zer}_p\tilde{\mathcal{A}}$. *One inserts for uncoupled* i, k

1. $i \frac{\quad}{} k$ *for every* $i \frac{\quad}{} \not\!\emptyset \frac{\quad}{} k$ *with* $\not\!\emptyset \in q$; $i, k \in F$ ($\text{zer}_q\mathcal{S}^{FF.O}$),
2. $i \longleftarrow k$ *for every* $i \longleftarrow \not\!\emptyset \longleftarrow k$ *with* $\not\!\emptyset \in p$; $i \in u$; $k \in F$, ($\mathcal{D}_{uF}$),
   and for every $i \longleftarrow \not\!\emptyset \frac{\quad}{} k$ *with* $\not\!\emptyset \in q$; $i \in u, k \in v$ ($\mathcal{D}_{uq}\mathcal{P}_{q|v.C}$).
   Next, one adds for any $i, k \in u$
3. i - - - k *for every* $i \longleftarrow \not\!\emptyset \longrightarrow k$ *with* $\not\!\emptyset \in p$ ($\mathcal{D}_{up}\mathcal{D}^{\mathrm{T}}_{up}$),
   and for every $i \longleftarrow \not\!\emptyset \frac{\quad}{} \not\!\emptyset \longrightarrow k$ *with* $\not\!\emptyset \in q$ ($\mathcal{D}_{uq}\mathcal{S}_{qq|vC}\mathcal{D}^{\mathrm{T}}_{up}$).

Then one keeps the subgraph induced by nodes $N = (u, v)$.

COROLLARY 9. Generating $G^{V\setminus\{\emptyset,M\}}_{\text{sum}}$ from G^V_{par}. *We let* $V = \{M, \bar{M}\}$, $\mathcal{B} = \text{inv}_M\tilde{\mathcal{A}}$. *One inserts for uncoupled* $i, k \in V$

1. $i \longleftarrow k$ *for every* $i \longleftarrow \not\!\emptyset \longleftarrow k$ *with* $\not\!\emptyset \in M$ ($\mathcal{B}$).
   Next, one adds for any $i, k \in \bar{M}$
2. i - - - k *for every* $i \longleftarrow \not\!\emptyset \longrightarrow k$ *with* $\not\!\emptyset \in M$ ($\mathcal{B}_{\bar{M}M}\mathcal{B}^{\mathrm{T}}_{\bar{M}M}$).

Then, one removes edges and nodes of M.

We note that in Section 3, M had been split for Corollary 5 into ordered components as $M = (p, q)$ with $p \subset O$, $q \subset F$, so that $\mathcal{B} = \text{zer}_q\mathcal{D} = \text{zer}_p\mathcal{E}$.

COROLLARY 10. Generating $G^{V\setminus\{C,M\}}_{\text{sum}}$ from $G^{V\setminus\{\emptyset,M\}}_{\text{sum}}$. *We let again* $\bar{M} = (u, w)$, $w = \{v, C\}$, $\mathcal{Q}_{\bar{M}\bar{M}} = \text{zer}_w\text{In}[\mathcal{I}_{\bar{M}\bar{M}} + \mathcal{B}_{\bar{M}M}\mathcal{B}^{\mathrm{T}}_{\bar{M}M}]$. *One inserts for uncoupled* i, k

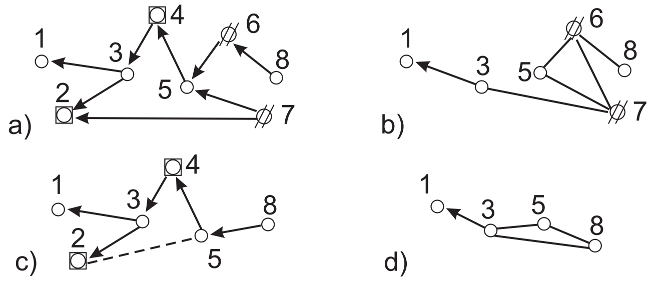
1. *i - - - k for every i - - - $\boxed{\circ}$ - - - k with $\boxed{\circ} \in w$; $i, k \in \bar{M}$ $(\mathcal{Q}_{\bar{M}\bar{M}})$;*
2. *$i \prec\!\!-\!\!-\, k$ for every i - - - $\boxed{\circ}$ $\prec\!\!-\!\!-\, k$ with $\boxed{\circ} \in w$; $i \in u$; $k \in w$ $(\mathcal{Q}_{uw}\mathcal{E}_{ww})$;*
3. *i —— k for every $i \,-\!\!-\!\!\succ$ $\boxed{\circ}$ - - - $\boxed{\circ}$ $\prec\!\!-\!\!-\, k$ with $\boxed{\circ} \in w$; $i, k \in v$ $(\mathcal{S}^{vv.Mu})$.*

Then, one replaces every edge within v by a full line, every i - - - k between u and v by $i \prec\!\!-\!\!-\, k \subset v$; and keeps the subgraph induced by nodes $N = (u, v)$.

Thus in general, parent graphs are not closed under conditioning nor under marginalising since an additional type of edge is needed to represent the resulting independence structure, see Proposition 4. Also, $G^{V\setminus\{\emptyset,M\}}_{\text{sum}}$ is not closed under conditioning and $G^{V\setminus\{C,\emptyset\}}_{\text{sum}}$ not under marginalising.

EXAMPLE 2. Path constructions of G^{V}_{sum} for $M = q$ and $p = \emptyset$. *The node set of the parent graph is $V = (1, \dots, 8)$. 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 2. *a) The generating graph G^{V}_{par}, b) $G^{V\setminus\{C,\emptyset\}}_{\text{sum}}$, c) $G^{V\setminus\{\emptyset,M\}}_{\text{sum}}$, d) G^{N}_{sum}.*



In this example, the summary graph model is equivalent to a triangular system in $N = (1, 3, 5, 8)$ even though $G^{V\setminus\{\emptyset,M\}}_{\text{sum}}$ is not equivalent to any directed acyclic graph since it contains the chordless pure-collision path $3 \,-\!\!-\!\!\succ 2$ - - - $5 \prec\!\!-\!\!-\, 8$.

5.3. *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 11. Generating $G_{\text{sum}}^{V\setminus\{C,m\}}$ from $G_{\text{sum}}^{V\setminus\{c,m\}}$. *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. $i \text{ - - - } k$ *for every* $i \text{ - - - } \boxed{\circ} \text{ - - - } k$ *with* $\boxed{\circ} \in r$; $i, k \in \mu$ $(\mathcal{Q}_{\mu\mu})$;
2. $i \prec\!\!\text{—}\, k$ *for every* $i \text{ - - - } \boxed{\circ} \prec\!\!\text{—}\, k$ *with* $\boxed{\circ} \in r$; $i \in o$; $k \in \psi$ $(\mathcal{Q}_{or}\mathcal{B}_{r\psi})$;
3. $i \text{ —— } k$ *for every* $i \text{—}\!\!\succ \boxed{\circ} \text{ - - - } \boxed{\circ} \prec\!\!\text{—}\, k$ *with* $\boxed{\circ} \in r$; $i, k \in \psi$ $(\mathcal{S}^{\psi\psi.mo})$.

Then, one replaces all $i \text{ - - - } k$ *present between* o *and* r *by* $i \prec\!\!\text{—}\, k \in r$, *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 12. Generating $G_{\text{sum}}^{V\setminus\{C,M\}}$ from $G_{\text{sum}}^{V\setminus\{C,m\}}$. *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* $\mathcal{H}''$, $\mathcal{K} = \text{zer}_M \mathcal{H}''$, *and insert for uncoupled nodes* i, k

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

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

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

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

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, which had been stated but not proven in Cox and Wermuth (1996). For larger classes of general densities than summary graphs, it was proven for MC-graphs by Koster (2002) and for another class by Richardson (2003).

Lemma 4. Sadeghi (2008). A path criterion for the global Markov property *A summary graph,* G_{sum}^N *implies* $\alpha \perp\!\!\!\perp \beta | C\gamma$ *if and only if it contains no path from a node* $i \in \alpha$ *to a node* $k \in \beta$ *such that for its inner nodes*
(i) *every collision node is in* γ *or is an ancestor of* γ *and*
(ii) *every other node is outside* γ.

6. Discussion. We have studied in this paper distributions that are generated over a parent graph in node set V. For this process, we assumed (1) a full ordering of V, (2) that the independence structure of f_V is captured by G^V_{par} and (3) that each ik-arrow present in G^V_{par} corresponds to a non-vanishing dependence of Y_i on Y_k conditionally given $Y_{\text{par}_i \setminus k}$. For the purpose of studying consequences of a given substantive research hypothesis, these are weak assumptions.

The reason is, for any uncoupled node pair (i, k) and one of the following two subgraphs

$$i \prec\!\!\!-\!\!\!- l \,-\!\!\!-\!\!\!\succ k, \qquad i \,-\!\!\!-\!\!\!\succ h \prec\!\!\!-\!\!\!- k$$

present in G^V_{par}, we know that the densities considered are association-inducing. This means, for c another subset of V, and $i \perp\!\!\!\perp k | lc$ specified by the first subgraph, that there is a non-vanishing association for Y_i and Y_k given Y_c and for $i \perp\!\!\!\perp k | c$ specified by the second subgraph, that there is a non-vanishing association for Y_i and Y_k given Y_c and Y_h, see Corollary 3 of Wermuth and Cox (2004).

The excluded class of distributions, in which both types of independence statements hold simultaneously, are of limited interest for statistical modeling since they contain non-smooth models, such as the one in Example 7 of Bergsma and Rudas (2002).

Whenever one considers the associations induced after marginalising or conditioning, the effects of different types of paths may be important. Such effects can cancel and lead to independences that occur in addition to those implied by a generating parent graph, even for Gaussian distributions generated over G^V_{par}, see Wermuth and Cox (1998). Or, they may lead to distortions of generating dependences that can be so severe that qualitatively wrong conclusions are obtained if one is not alerted to such situations, see Wermuth and Cox (2008).

In a summary graph, the double edge, $i \overset{\prec\!\!-\!\!-}{- - -} k$, points to distortions due to direct confounding, a problem that is avoided for ik-dependences, whenever it is possible to randomly allocate individuals to the levels of Y_k and Y_i is a direct response of Y_k in the generating process and in G^V_{sum}.

In the absence of both direct confounding and other known causes for distortions, there may still be still so-called indirect confounding. It shows in a summary graph via a particular type of pure collision path, connecting nodes $i < k$, for $i \prec\!\!\!-\!\!\!- k$. These warning signals provided by summary graphs are essential, whenever one wants to understand the consequences of relatively complex research hypotheses captured by parent graphs or by multivariate regression graphs.

An MC-graph, see Koster (2002), contains even more of such warning signals, since there may be up to four types of edge coupling a given variable pair. Sadeghi (2008) obtains the MC-graph and proves its global Markov property by using the edge-inducing V-configurations in the following table and an active alternating path, as defined by Marchetti and Wermuth (2008).

Types of edge inducing V-configurations for MC-graphs by

	conditioning, $\boxed{\circ}$		marginalising, $\not\circ$	
	$\circ \leftarrow k$	$\circ \text{ - - - } k$	$\circ \rightarrow k$	$\circ \relbar\!\!\relbar\ k$
$i \rightarrow \circ$	$i \relbar\!\!\relbar k$	$i \rightarrow k$	$i \rightarrow k$	$i \relbar\!\!\relbar k$
$i \text{ - - - } \circ$	$\cdot$	$i \text{ - - - } k$	$i \rightarrow k$	$i \leftarrow k$
$i \leftarrow \circ$	$\cdot$	$\cdot$	$i \text{ - - - } k$	$i \leftarrow k$
$i \relbar\!\!\relbar \circ$	$\cdot$	$\cdot$	$\cdot$	$i \relbar\!\!\relbar k$

Furthermore, Sadeghi derives the summary graph as what he calls an edge-mininal MC-graph.

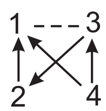
In a summary graph and in a MC-graph, there may be an ik-edge missing but no independence statement is implied for Y_i, Y_k so that they are, in general, no independence graphs even though both types capture the independence structure implied by the parent graph for $Y_{N|C}$. The simplest three-node graph of this type is

$$i \underset{\text{- - -}}{\leftarrow} j \leftarrow k, \tag{6.1}$$

One of the two involved V-configurations is closed by marginalising, the other by conditioning, see Corollaries 11 and 12. The graph is for the simplest type of a so-called instrumental variable model. In the linear case case, it is identifiable and the dependence represented by the arrow is a ratio of simple least-squares regression coefficients, $\beta_{i|k}/\beta_{j|k}$.

The small summary graph in Figure 3 is an independence graph and it alerts to the possible distortion of the generating dependence $1 \leftarrow 4$.

FIG 3. *A summary graph with* $2 \perp\!\!\!\perp 4|3$ *and distorting path* $1 \text{ - - - } 3 \leftarrow 4$ *for indirect confounding of dependence* $1 \leftarrow 4$ *and* $1 \leftarrow 2$ *undistorted by direct or indirect confounding.*



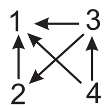
Models corresponding to many MC-graphs and to some summary graphs with double edges may be under-identified. But, given any summary graph, a corresponding triangular system can be derived for Y_u via equation (3.4). This may lead to a reduced model, see Cox and Wermuth (1990) with nonzero constraints, for which estimation is standard for a covering model.

Or, the independence equivalent maximal ancestral graph of Richardson and Spirtes (2002) can be derived.

Each maximal ancestral graph is an independence graph and, for Gaussian distributions, each corresponds to an identifiable model. These nice properties are obtained at the cost of no longer alerting to distortions of the types described above or to special types of parameter definition, such as the one in the above instrumental variable model.

The maximal ancestral graph corresponding to (6.1) is a complete directed acyclic graph in the order (i, j, k). The maximal ancestral graph to Figure 3 is given in Figure 4.

FIG 4. *The maximal ancestral graph that is independence equivalent to the graph in Fig. 3.*



Given the summary graph and linear relations, it is often possible to recover the generating dependence, see Wermuth and Cox (2008). In the example of Figures 4 and 3, the linear least-squares regression coefficient $\beta_{1|4.23}$, the coefficient corresponding to $1 \prec\!\!\!-\!\!\!- 4$, is corrected for indirect confounding that is due to the distorting path $1 \text{---} 3 \prec\!\!\!-\!\!\!- 4$.

In the above two small examples, the summary graph is independence equivalent to a directed acyclic graph, and in the Gaussian case, the corresponding models are parameter equivalent, i.e. the two sets of parameters are related by one-to-one transformations. From Proposition 4, we know when this will not be the case, but still lacking are both, an efficient algorithm to decide this in general and a positive characterization.

Acknowledgement. I thank D.R. Cox and G.M. Marchetti for their careful reading of the manuscript and insightful suggestions and the referees for their helpful comments.

REFERENCES

BERGSMA, W. AND RUDAS, T. (2002). Marginal models for categorical data. *Ann. Statist.* **30**, 140–159.

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. AND WERMUTH, N. (1990). An approximation to maximum-likelihood estimates in reduced models. *Biometrika* **77**, 747–761.

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

Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scan. J. Statist.* **23**, 115-116.

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., Dawid, A. P., Larsen, B. N. and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491–505.

Marchetti, G.M. and Wermuth, N. (2008). Matrix representations and independencies in directed acyclic graphs. *Ann. Statist.* **46**, to appear.

Paz, A. (2007). The annotated graph model for representing DAG-representable relations–algorithms approach. *Res. Rep., Technion, Haifa.*

Pearl, J. and Paz, A. (1987). A graph based logic for reasoning about relevance relations. In *Advances in Artificial Intelligence II* B. Duboulay, D. Hogg and L. Steels, (eds.) 357-363. North Holland: Amsterdam.

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*

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., Cox, D.R. and Marchetti, G.M. (2006). Covariance chains. *Bernoulli* **12**, 841-862.

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.

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