When Can Association Graphs Admit A Causal Interpretation?

(First Report)

Judea Pearl
Cognitive Systems Laboratory
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Nanny Wermuth
Psychologisches Institut
Gutenberg Universität, 6500 Mainz, Germany
wermuth@upymzc.sowi.uni-mainz.de

January 5, 1993

Summary

We discuss essentially linear structures which are adequately represented by association graphs called covariance graphs and concentration graphs. These do not explicitly indicate a process by which data could be generated in a stepwise fashion. Therefore, on their own, they do not suggest a causal interpretation. By contrast, each directed acyclic graph describes such a process and may offer a causal interpretation whenever this process is in agreement with substantive knowledge about causation among the variables under study. We derive conditions and procedures to decide for any given covariance graph or concentration graph whether all their pairwise independencies can be implied by some directed acyclic graph.

1 Introduction

Many of the most widely used statistical techniques center around the interpretation of the covariance matrix. The exponential family of the multivariate normal distribution provides a justification to reduce data to sample means, $\bar{y}_i$, $i = 1, ..., p$, and a covariance matrix, S, and for some purposes to sample means and a concentration matrix, S^{-1}, that is the inverse covariance matrix.

Under multivariate normality, a zero covariance, $\sigma_{ij} = 0$, means marginal independence and a zero concentration, $\sigma^{ij} = 0$, means conditional independence of variable pair Y_i, Y_j given all $p - 2$ remaining variables. In general, these conditions just indicate linear independencies, since the standardized versions of covariance and concentration are the marginal correlation coefficient, ρ_{ij}, and the partial correlation coefficient, $\rho_{ij.k}$, given all the remaining variables $k = \{1, ..., p\} \backslash \{i, j\}$:

$$\rho_{ij} = \sigma_{ij}(\sigma_{ii}\sigma_{jj})^{-1/2}, \rho_{ij.k} = -\sigma^{ij}(\sigma^{ii}\sigma^{jj})^{-1/2}.$$

Thus, data reduction to covariance or concentration matrix implies that essentially linear relations among the variables under study are of interest. These may be relevant not only for continuous but also for ordinally scaled variables or for dichotomous variables. But, since this is a fairly strong assumption in some applications, checks for nonlinearity or special linearisatious are needed, such as those discussed in [4, 5] and in [18].

A covariance matrix with some zero off-diagonal entries can be represented by an undirected graph (of dashed edges) with some edges missing [6] called **covariance graph**, abbreviated by G_{cov} here. Similarly, a concentration matrix with some zero entries can be represented by an undirected graph (of full lines) with some edges missing [12] called **concentration graph**, abbreviated by G_{con} here.

A **complete** covariance or concentration graph has no missing edges, that is, it has exactly one edge for each variable pair, $1/2\ p\ (p-1)$ in all. There can be a number of reasons for considering special incomplete undirected graphs. For instance, it could be that

- the linear independencies expressed thereby are themselves of substantive interest,
- reduction of dimensionality is needed to keep a reasonable balance between the number of parameters to be estimated and the number of available independent observations, or
- it is desired to decide whether the structure given by a covariance or concentration graph could have been generated by some stepwise process of univariate recursive regressions.

We concentrate here on the last aspect. This implies in particular that we consider covariance and concentrations graphs less as representing models of their own standing but, rather, a possible reflection of some directed acyclic graph, abbreviated as **dag**, which represents a recursive process for stepwise data generation. Such a process would admit a causal interpretation if the order of the variables involved in the process, especially their classification into response and explanatory variables, is in agreement with subject-matter knowledge about the causal process [3].

Each dag can be thought of as specifying dependence relationships between an ordered set of variables $\{Y_1, ..., Y_p\}$ where $j > i$ designates Y_j as potentially explanatory for Y_i; it becomes actively explanatory in case an arrow is drawn from Y_j to Y_i. A missing edge for variable pair Y_i, Y_j has then the interpretation: response Y_i is (linearly) independent of Y_j given the remaining potential explanatory variables of Y_i, $\{Y_{i+1} ..., Y_{j-1}, Y_{j+1}, ..., Y_p\}$. The set of missing edges defines the independence structure of the dag and permits[1] the stronger statement that Y_i is conditionally independent of all variables in $\{Y_{i+1}, ..., Y_p\}$ that are not linked to Y_i, given those that are linked to Y_i. Each dag **implies** a set of pairwise marginal (linear) independencies, i.e. a covariance graph, and a set of pairwise conditional (linear) independencies given all $p-2$ variables, i.e. a concentration graph. By "implied"

[1] Provided S is non-singular.

we mean independencies that must hold in every distribution that fulfills all the missing-link conditions shown in the dag. This excludes independencies that are just introduced by equalities among the numerical values of some parameters in the model. From a given dag, these implied independencies may be read off directly, using the graphical criterion of d-separation [13, page 117] and [10].

For example the following Markov chain

$$\begin{array}{ccccc} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 \\ \circ \longleftarrow & \circ \longleftarrow & \circ \longleftarrow & \circ \longleftarrow & \circ \end{array}$$

is a dag which implies a complete covariance graph, i.e. no marginal independencies, and a chain-like concentration graph, obtained by removing the arrow heads. If r, s, t, u denote correlations between pairs $(Y_1, Y_2), (Y_2, Y_3), (Y_3, Y_4), (Y_4, Y_5)$, then the correlation matrix implied by the Markov chain has the form

$$\begin{array}{ccccc} 1 & r & rs & rst & rstu \\ \bullet & 1 & s & st & stu \\ \bullet & \bullet & 1 & t & tu \\ \bullet & \bullet & \bullet & 1 & u \\ \bullet & \bullet & \bullet & \bullet & 1. \end{array}$$

It can be used to explain why it may be difficult to decide on marginal linear independence merely from observed marginal correlations. Suppose $r = s = t = u$ and $r = 0.3$ then $r^2 = 0.09.$, $r^3 = 0.027$, $r^4 = 0.0081$ would be close to zero; even when $r = 0.6$, $r^4 = 0.13$ would still not be far from zero. Thus, if we know that the data are generated by the above Markov chain we can conclude that all marginal correlation are nonzero but a correlation is weaker the further the variables are apart in the chain. However, if the process by which the data actually are generated is not known, and we have to rely merely on data inspection or associated tests, we may judge some of the smaller correlations to be zero apart from sample fluctuations i.e. we come to erroneous conclusions about the covariance structure represented by the covariance graph. Similar arguments apply to the concentration graph.

Formal tests for agreement of an observed covariance matrix or concentration matrix with a corresponding hypothesized association graph are available. An arbitrary pattern of zeros in a covariance matrix is a special case of hypotheses linear in covariances [1] and an arbitrary pattern of zeros in a concentration matrix is a covariance selection model [7]. Maximum likelihood estimates and associated likelihood ratio tests are available under the assumption of multivariate normality and may for instance be computed with the help of LISREL [11] for the former and with the help of MIM [9] for the latter. The same estimates are called quasi-likelihood estimates if just the assumption of linearity is retained.

The problem of finding a causal explanation for a complete list of *all* conditional and marginal independencies was treated in [16], and was shown to require a polynomial number of steps [8]. However, the length of this list can be very large indeed, since enumerating all conditional independencies might require the processing of an exponential number of submatrices (i.e., all the majors of G_{cov}). The current paper attempts to find a causal interpretation on the basis of a more limited information, assuming that the only vanishing dependencies available to the analyst are those corresponding to the zero entries of the covariance and concentration matrices.

To this end, we pose the following set of problems:

• Given a covariance graph, G_{cov}, decide whether it could have been generated by a directed acyclic graph

(i) with the same nodes and edges,

(ii) with the same nodes but fewer edges,

(iii) with some additional nodes (representing latent or hidden variables).

• Given a concentration graph, G_{con}, decide whether it could have been generated by a directed acyclic graph

(iv) with the same nodes and edges,

(v) with the same nodes but fewer edges,

(vi) with additional (latent) nodes.

• Given a covariance and a concentration graphs, decide whether

(vii) both could have been generated by the same directed acyclic graph.

Theorems 1 through 7 of the next section present declarative and procedural solutions to these seven problems, respectively. Proofs can be found in a forthcoming paper [14].

2 Main Results

Definition 1 *A directed acyclic graph D is said to imply an independence I if I holds in every probability P that fulfills all the missing-link conditions of D, that is, choosing any variable ordering consistent with the direction of arrows in D, P renders each variable Y_i independent on its inactive explanatory variables (nonadjacent predecessors), given its active explanatory variables (adjacent predecessors). If D fails to imply an independence for pair (i, j) we say that D implies a dependence for that pair.*

Lemma 1 *D implies an independence I iff I satisfies the d-separation criterion in D* [13]. *In particular, D implies a marginal independency $\sigma_{ij} = 0$ if every path between nodes i and j in D traverses at least one pair of converging arrows (i.e., $\longrightarrow \star \longleftarrow$). Likewise, D implies a conditional independence given all remaining variables, $\sigma^{ij} = 0$, if every path between nodes i and j in D traverses at least one pair of non-converging arrows (i.e., either $\longrightarrow \star \longrightarrow$ or $\longleftarrow \star \longrightarrow$).*

Definition 2 (*generate and coincide*) *A dag D is said to (a) generate a covariance graph G_{cov} if the set of marginal independencies implied by D matches the set of missing edges of G_{cov} and (b) to coincide with G_{cov} if D not only generates G_{cov} but also has the same edges as G_{cov}.*

Similarly, a dag D is said to (a) *generate* a concentration graph G_{con} if the set of pairwise conditional independencies given all remaining variables implied by D matches the set of missing edges of G_{con} and (b) to *coincide* with a G_{con} if D not only generates G_{con} but also has the same edges as G_{con}.

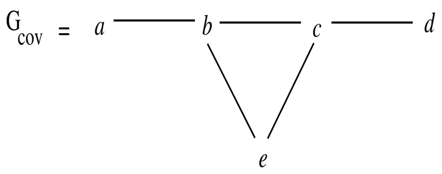
Definition 3 ($\vee$-*configuration*) *A triplet of nodes (a, b, c) in an undirected graph G is said to be a $\vee$-configuration (pronounced as vee-configuration), if ab and bc are edges in G while ac is not.*

Definition 4 (*sink orientation*) *Given an undirected graph G, the sink orientation of G results from assigning arrows $a \longrightarrow b \longleftarrow c$ to each $\vee$-configuration (a, b, c) in G.*

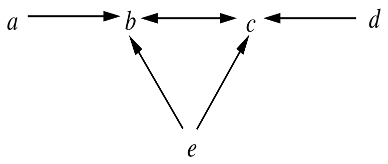
Note that a sink orientation may contain bi-directed edges, since two opposing arrows may be assigned to the same edge. For example, the sink orientation of the 4-chain a—b—c—d is given by $a \longrightarrow b \longleftrightarrow c \longleftarrow d$, since both (a, b, c) and (b, c, d) are $\vee$-configurations. We mark that the sink orientation of a graph with n nodes can be formed in $O(n^3)$ steps.

Theorem 1 *The following statements are equivalent. A covariance graph G_{cov} coincides with a dag D if and only if (i) no edge in the sink orientation of G_{cov} is bi-directed; (ii) G_{cov} contains no chordless 4-chain; Moreover, if the sink orientation of G_{cov} contains no bi-directed edges, then every acyclic orientation of the remaining undirected edges constitutes a dag that coincides with G_{cov}.*

Example 1 Consider the graph



The 4-chain a—b—c—d is chordless, hence, there is no dag that coincides with G_{cov}. Indeed, the $\vee$-orientation of G_{cov} is



which renders the edge $b \longleftrightarrow c$ bi-directed.

Turning to problem (ii), we first define the notions of **sink completion** and **exterior cliques**.

Definition 5 (*sink completion*) *A sink completion of an undirected graph G_{cov} is any dag obtained from G_{cov} as follows: a) the sink orientation of G_{cov} is formed, b) all bi-directed edges are removed, c) the remaining undirected edges are oriented to form a dag.*

Lemma 2 *Every undirected graph has at least one sink completion.*

This follows from the facts that (1) every sink orientation is acyclic and (2) any acyclic partial orientation of a graph can be completed to form a dag.

Definition 6 (*exterior cliques*) *A clique in an undirected graph is said to be exterior if it contains at least one node that is adjacent only to nodes in that clique. We call such a node "extremal".*

Note that finding an extremal node in a graph with n nodes and E edges takes at most nE^2 steps; for each node i we test whether every two neighbors of i are adjacent to each other.

Theorem 2 *The following statements are equivalent: (i) G_{cov} is generated by a directed acyclic graph; (ii) a sink completion of G_{cov} implies all edges of G_{cov}; (iii) every edge of G_{cov} resides in an exterior clique.*

Example 2 Consider again the graph G_{cov} of Example 1. The sink completion of G_{cov} is given by the dag $D = a \longrightarrow b \longleftarrow e \longrightarrow c \longleftarrow d$ which implies all edges of G_{cov}. In particular, the edge b—c, missing from D, is implied by the path $b \longleftarrow e \longrightarrow c$, which does not contain converging arrows (see Lemma 1.) Hence D generates G_{cov}. Indeed, following condition (iii), the exterior cliques of G_{cov} are (a, b), (b, c, d) and (c, d), which contain every edge of G_{cov}.

A simple example of a graph G_{cov} that cannot be generated by a dag is the 4-chain a—b—c—d. The sink completion of G_{cov} is the disconnected dag $a \longrightarrow b \quad c \longleftarrow d$ which does not imply the edge b—c in G_{cov}. Likewise, G_{cov} has only two exterior cliques, (a, b) and (c, d), thus violating condition (iii) relative to edge b—c.

Turning to Problem (iii), we have:

Theorem 3 *Every covariance graph G_{cov} can be generated by a directed acyclic graph with additional nodes by by replacing each edge ab of G_{cov} with $a \longleftarrow (*) \longrightarrow b$, where $(*)$ represents an unobserved variable.*

Remark: To reduce the number of latent nodes, it is not necessary to replace each edge of G_{cov}, but only those which are not implied by the sink completions of G_{cov}.

Example 3 The 4-chain a—b—c—d is generated by the dag $a \longrightarrow b \longleftarrow (*) \longrightarrow c \longleftarrow d$, where the replaced edge bc is the only one that is not implied by the sink completion of this chain (see Example 2). We now turn to problems (iv) - (vi) where we seek a dag that agrees with the structure of a given concentration matrix.

Theorem 4 *A concentration graph G_{con} coincides with some dag D if and only if G_{con} is a chordal graph, i.e. it contains no chordless n-cycle, $n > 3$. (Effective tests for chordality are well known in the literature* [15].)

Theorem 5 *A concentration graph G_{con} can be generated by some dag if the following procedure removes all edges of G_{con}:*

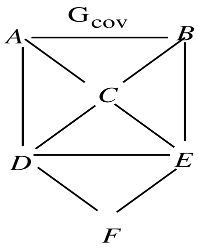
1. *Find an exterior clique, remove all its extremal nodes and their associated edges, and mark all remaining edges in that clique.*

2. *Repeat the process on the reduced graph until either all nodes have been eliminated or, in case the subgraph contains no exterior clique, choose any marked edge and remove it from the graph.*

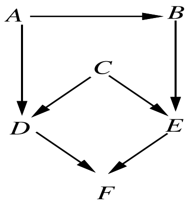
3. *Repeat step 1 and 2.*

Conversely, if there is no removal ordering that leads to an empty graph then G_{con} cannot be generated by a dag.

Example 4 To see how the marking procedure helps identify a generating dag, consider the graph G_{con} below.

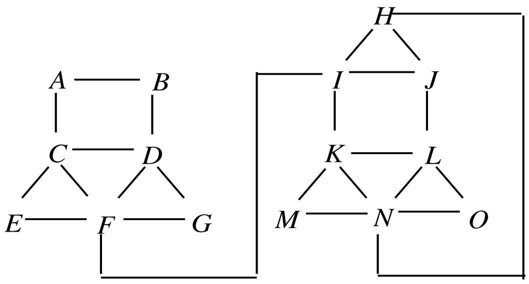


The cliques are: ABC, ACD, BCE, CDE and DEF. Only DEF is exterior, with extremal vertex F. Removing F together with edges DF and FE, leaves a subgraph with no exterior clique. However, edge DE is marked, so it can be removed, which creates new exterior cliques, allowing the process to continue till all edges are removed. Indeed, directing arrows toward each node removed in step 2, yields the following dag which generates G_{con}



The following example (due to Verma) shows that the condition provided by the elimination procedure of Theorem 5, though sufficient, is not necessary. That is, failure to eliminate all edges in one ordering does not imply that no elimination ordering exists.

Consider the following concentration graph: The only extremal vertices are E, G, M, and

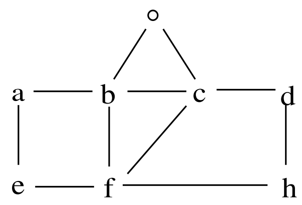


O. If we remove these in any order it will result in a graph with no exterior cliques and with the following marked links: CF, DF, KN, and LN. If we first remove the marked edges FC and FD, the process will halt (because the cycle $A-B-C-D-A$ cannot be eliminated.) However, if we first remove the marked edges LN and KN, then we will find a good elimination ordering: ..., N, H, J, L, K, I, F, ...

The fact that it is impossible by local means to decide which of the marked edges should be removed first renders the decision problem a difficult one. While in the example above it is clear that one should postpone the removal of FC and FD, because it leads to an impasse (the cycle $A - B - C - D - A$), such local clues are not available in the general case. Indeed, Verma and Pearl have shown that the problem is NP-Complete [17]. Nevertheless, effective necessary conditions are available, which makes the decision problem feasible when the number of nodes is not too large.

Lemma 3 *The following are necessary conditions for a concentration graph G_{con} to be generated by a dag. (1) every chordless n-cycle of G_{con}, $n > 3$, must have at least one edge that resides in some k-clique, $k > 2$. (2) G_{con} must have at least one exterior clique.*

Example 5 The graph G_{con} below shows that these two conditions are not sufficient; it satisfies both, yet it cannot be generated by any dag.



Theorem 6 *A concentration graph G_{con} can be generated by a dag with additional latent nodes if and only if it can be generated by a dag without latent nodes.*

In other words, latent variables do not improve the power of dags to generate richer patterns of vanishing concentrations. This is in sharp contrast to patterns of vanishing covariances (Theorem 3); every such pattern can be generated by a dag with latent nodes.

Finally, turning to problem (vii) we have:

Theorem 7 *A covariance graph G_{cov} and a concentration graph G_{con} can both be generated by the same dag only if the following three conditions hold:*

1. *G_{cov} and G_{con} must each be generated by a dag.*

2. *The set of nodes of every exterior clique of G_{cov} must induce a subgraph of G_{con} that can be generated by a dag.*

3. *Every edge of G_{con} that does not reside in a larger clique must also be an edge of G_{cov}.*

We are not sure at this point whether the conditions above are sufficient; so far we were not able to find a pair of undirected graphs that fulfill these four conditions and yet cannot be generated by a dag.

3 Conclusions

This paper provides conditions and procedures for deciding if patterns of independencies found in covariance and concentration matrices can be generated by a stepwise recursive process represented by some directed acyclic graph. If such an agreement is found, we know that one or several causal processes could be responsible for the observed independencies, and our procedures could then be used to elucidate the graphical structure common to these processes, so as to evaluate their compatibility against substantive knowledge of the domain.

If we find that the observed pattern of independencies does not agree with any stepwise recursive process, then there are a number of different possibilities. For instance,
– some weak dependencies could have been mistaken for independencies and led to the wrong omission of edges from the covariance or concentration graphs.
– some of the observed linear dependencies reflect accidental cancellations or hide actual nonlinear relations, or
– the process responsible for the data is non-recursive, involving aggregated variables, simultenous reciprocal interactions, or mixtures of several causal processes.

In order to recognize accidental independencies it would be helpful to conduct several longitudinal studies under slightly varying conditions. In such studies the covariances for the same set of variables is estimated under different conditions and the variations in the conditions would typically affect the numerical values of the parameters. But, if the data were generated by a causal process represented by some directed acyclic graph, then the basic structural properties reflected in the missing edges of that graph should remain unchanged. Under such assumptions, the pattern of independencies that is "implied" by the dag (see Definition 1) would persist despite the fluctuating conditions, while accidental independencies would be perturbed and hence discarded. We regard this as a possibility to pursue Cochran's suggestion [2, page 252] that "when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible and plan observational studies to discover whether each of these consequences is found to hold."

References

[1] Anderson, T.W., "Asymptotically Efficient Estimation of Covariance Matrices With Linear Structure," *Ann. Statist.*, 1, 135-141, 1973.

[2] Cochran, W.G., "The Planning of Observational Studies of Human Populations," *J. Roy. Statist. Soc. A*, 128:234-265, 1965.

[3] Cox, D.R., "Causality: Some Statistical Aspects," *J. Roy. Statist. Soc.* Ser. A., 155, 291-301, 1992.

[4] Cox, D.R. and Wermuth, N., "Tests of Linearity, Multivariate Normality and the Adequacy of Linear Scores," Berichte zur Stochastik und verw. Gebiete, Universität Mainz, 92-6, 1992a.

[5] Cox, D.R. and Wermuth, N., "Response Models for Mixed Binary and Quantitative Variables," *Biometrika*, 79, 441-461, 1992b.

[6] Cox, D.R. and Wermuth, N., "Linear Dependencies Represented by Chain Graphs," Berichte zur Stochastik und verw. Gebiete, Universität Mainz, 92-3, 1992c.

[7] Dempster, A.P., "Covariance Selection," *Biometrics*, 28, 157-175, 1972.

[8] Dori, D. and Tarsi, M., "A Simple Algorithm to Construct a Consistent Extension of a Partially Oriented Graph," Computer Science Department, Tel-Aviv University. Also *Technical Report R-185*, UCLA, Cognitive Systems Laboratory, October 1992.

[9] Edwards, D., "Graphical Modelling With MIM," Manual.

[10] Geiger, D., Verma, T.S., and Pearl, J., "Identifying Independence in Bayesian Networks," *Networks*, 20, John Wiley and Sons, Sussex, England, 507-534, 1990.

[11] Jöreskog, D.G. and Sörbom, D., *LISREL IV, Analysis of Linear Structural Relationships by Maximum-Likelihood, Instrumental Variables and Least Squares Methods*, 3rd Edition, Mooresville: Scientific Software, 1978.

[12] Kiiveri, H. and Speed, T.P., "Structural Analysis of Multivariate Data: A Review," in S. Leinhardt (Ed.), *Sociological Methodology*, San Francisco: Jossey Bass, 209-289, 1982.

[13] Pearl, J., *Probabilistic Reasoning in Intelligence Systems*, Morgan Kaufmann, San Mateo, CA, 1988.

[14] Pearl, J. and Wermuth, N., "When Can Association Graphs Admit a Causal Interpretation?," UCLA, Cognitive Systems Laboratory, *Technical Report R-183-L*, November 1992.

[15] Tarjan, R.E. and Yannakakis, M., "Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs and Selectively Reduce Acyclic Hypergraphs," *SIAM Journal of Computing*, 13, 566-579, 1984.

[16] Verma, T. S., and Pearl, J., "An Algorithm for Deciding if a Set of Observed Independencies has a Causal Explanation," in *Proceedings, 8th Conference on Uncertainty in Artificial Intelligence*, Stanford, CA, Morgan Kaufmann, San Mateo, CA, 323-330, July 1992.

[17] Verma, T.S. and Pearl, J., "Deciding Morality of Graphs is NP-complete," *Technical Report R-188*, UCLA, Cognitive Systems Laboratory, October 1992.

[18] Wermuth, N. and Cox, D.R., "On the Relation Between Interactions Obtained With Alternative Codings of Discrete Variables," *Methodika*, VI, 76-85, 1992.