Ann Inst Stat Math (2014) 66:345–367
DOI 10.1007/s10463-013-0417-x

Qualitative inequalities for squared partial correlations of a Gaussian random vector

Sanjay Chaudhuri

Received: 4 January 2011 / Revised: 5 February 2013 / Published online: 14 August 2013

Abstract We describe various sets of conditional independence relationships, sufficient for qualitatively comparing non-vanishing squared partial correlations of a Gaussian random vector. These sufficient conditions are satisfied by several graphical Markov models. Rules for comparing degree of association among the vertices of such Gaussian graphical models are also developed. We apply these rules to compare conditional dependencies on Gaussian trees. In particular for trees, we show that such dependence can be completely characterised by the length of the paths joining the dependent vertices to each other and to the vertices conditioned on. We also apply our results to postulate rules for model selection for polytree models. Our rules apply to mutual information of Gaussian random vectors as well.

Keywords Inequalities · Graphical Markov models · Mutual information · Squared partial correlation · Tree models

1 Introduction

In graphical Markov models literature, several attempts have been made to characterise the degree of conditional association among the vertices by the structure of

This research was partially supported by Grant R-155-000-081-112 from National University of Singapore.

Electronic supplementary material The online version of this article (doi:10.1007/s10463-013-0417-x) contains supplementary material, which is available to authorized users.

S. Chaudhuri (✉)
Department of Statistics and Applied Probability, National University of Singapore,
Singapore117546, Singapore
e-mail: sanjay@stat.nus.edu.sg

the underlying graph. Such knowledge is considered useful in model selection. For example, Cheng et al. (2002) describe an algorithm of model selection for directed acyclic graphs (DAG) which assumes that the mutual information has a monotone relationship with certain structure based length of the path. Examples (Chickering and Meek 2006) show that such a *monotone DAG faithfulness* property or a similar *compound monotone DAG faithfulness* property do not hold even for simple binary DAGs. In fact, except in some specific cases e.g. Greenland (2003) in epidemiology, Spirtes et al. (2000, causal pipes) in causal analysis, no result is known in this context.

A more general problem is to order the squared partial correlation coefficients among the components of a Gaussian random vector. For these random vectors, squared partial correlation coefficients completely measure the degree of association between its components conditional on a subset of the components. This measure is a polynomial in the entries of their covariance matrices. Thus in many situations it is beneficial to be able to order squared partial correlation coefficients in a way, such that the ordering does not depend on the specific values of the covariances.

Simple counter-examples show that such *qualitative* comparisons cannot hold unless the covariance matrix belongs to certain subsets of positive definite matrices. In this article, we specify such subsets by conditional independence relationships. For a graphical Markov model validity of such relationships can be simply read off from the underlying graph. Thus rules for comparing degree of association on various Gaussian graphical models can be developed.

In this article we show that, certain conditional independence relationships holding, suitable squared partial correlations can be qualitatively compared. We make two kinds of comparisons. In the first, the set of components conditioned on (conditionate) are kept fixed and we change the dependent vertices (correlates). More importantly, in the second, we fix the two correlates and compare their degree of dependence by varying the conditionates. The sufficient conditional independence relationships are satisfied by several graphical Markov models. Using relevant *separation* criteria (e.g. separation for undirected graphs (UG) (see Definition 1), d-separation for DAGs (Verma and Pearl 1990) (see Definition 4), m-separation for mixed ancestral graphs (MAGs) (see supplement) (Richardson and Spirtes 2002) etc., we postulate sufficient structural conditions for comparing conditional association on them. We emphasize that the specific graphical Markov models are used as illustrations. Our results apply to a much wider class of models. Furthermore, using the fact that for tree and polytree (DAGs without any undirected cycles either or singly connected directed acyclic graphs) models, any two connected components have exactly one path joining them, these structural criteria can be simplified to path based rules for comparison. We discuss such rules for trees in details, where it is also shown that our rules for comparing the squared partial correlations are complete.

The inequalities discussed here have theoretical interest as new properties of Gaussian random vectors and directly translate to corresponding conditional non-Shannon type information inequalities (Zhang and Yeung 1997; Matúš 2006, 2007). Matúš (2005) considers implications of one set of conditional independence relations on other conditional independencies for Gaussian random vectors. Furthermore, he describes a way to determine such implications using the ring of polynomials generated by the entries of the correlation matrices with some additional indeterminates. Our results describe some polynomial inequalities these rings satisfy.

Our main motivation comes from the Gaussian graphical Markov models. These results are canonical and sufficient to postulate structure based rules to order dependencies on several of them. We improve upon Chaudhuri and Richardson (2003), Chaudhuri (2005), who only consider polytree models. These results can be used in determining the distortion effects (Wermuth and Cox 2008) and monotonic effects (VanderWeele and Robins 2007, 2010) of confounded variables in epidemiology and causal network analysis (see also Greenland and Pearl 2011). We postulate necessary and sufficient conditions for determining structures on a class of polytree models. These conditions can be directly applied in model selection, specially in mapping river flow and drainage networks where such polytree models occur naturally (Rodríguez-Iturbe and Rinaldo 2001). In real data analysis, these inequalities would be useful for model selection, specially among various graphical Markov models (Cheng et al. 2002; Shimizu et al. 2006). For these models our results would translate to hypothesis connected to the structure of the graph. These hypothesis can be tested from the observed data. Structure based inequalities may also be used as constraints in estimation with missing values. They are also relevant in choosing prior distributions in Bayesian procedures. The qualitative bounds can be used in selecting stratifying variables in designing surveys, gathering most relevant information in forensic sciences and building strategies for constrained searches. Further, these results may have applications in designing effective updating and blocking strategies in Gibbs sampling and Markov chain monte carlo procedures (see e.g. Roberts and Sahu 1997 etc).

2 Squared partial correlation inequalities

Suppose $V \sim N(\mu, \Sigma)$ with a positive definite Σ. Let a, b, c, c', z, z', x etc. be the components and B, Z etc. be the subsets of components of V. In this article V will also denote the vertex set of the underlying graph (see supplement for more details). Let $\emptyset$ denote the empty set.

The squared partial correlation coefficient ($\rho^2_{ac|Z}$) between a and c conditional on Z is defined by:

$$\rho^2_{ac|Z} = \frac{\left(\sigma_{ac} - \Sigma_{aZ}\Sigma^{-1}_{ZZ}\sigma_{cZ}\right)^2}{\left(\sigma_{aa} - \Sigma_{aZ}\Sigma^{-1}_{ZZ}\sigma_{aZ}\right)\left(\sigma_{cc} - \Sigma_{cZ}\Sigma^{-1}_{ZZ}\sigma_{cZ}\right)} = 1 - \mathrm{e}^{-2\mathrm{Inf}(a \perp\!\!\!\perp c|Z)}. \quad (1)$$

Here σ_{ab} and Σ_{aZ} respectively denote the (a, b)th element and $a \times Z$ submatrix of Σ. $\mathrm{Inf}(a \perp\!\!\!\perp c|Z)$ is the mutual information (Whittaker 2008, information proper) of a and c given Z. From (1) it follows that the mutual information is a monotone increasing function of the corresponding squared partial correlation. Thus the qualitative inequalities for $\rho^2_{ac|Z}$ presented below applies to $\mathrm{Inf}(a \perp\!\!\!\perp c|Z)$ as well.

2.1 Comparing conditional dependence with a fixed conditionate

We first fix a subset Z to be conditioned and one correlate a. The squared partial correlation is compared by changing the other correlate from c to c'.

Theorem 1 *Suppose $c' \perp\!\!\!\perp a|cZ$, then $\rho^2_{ac'|Z} \le \rho^2_{ac|Z}$.*

Theorem 1 is a conditional version of the well-known *information inequality* (Cover and Thomas 2006) and holds in general for mutual information of any distribution. For graphical Markov models the condition holds if c' is separated from a given c and Z. Further, for trees the condition is satisfied if c lies on the path joining a and c'. Thus longer path implies weaker dependence in this case.

For polytree models the condition depends on the arrangement of the arrows on the path joining a, c and c'. The condition is satisfied if two arrowheads do not meet at c on the path joining a and c', (i.e. c is not a *collider* on the path joining a and c', see Definition 3). As for example, in Fig. 1a with $Z = \{z_1, z_2, z_3, z_4\}$, using the d-separation criterion (see Definition 4) we get, $c_3 \perp\!\!\!\perp a|Zc_2$. Theorem 1 ensures that $\rho^2_{ac_3|Z} \le \rho^2_{ac_2|Z}$. The same d-separation criterion however implies that $c_3 \not\perp\!\!\!\perp a|Zc_1$, so there is no guaranty the $\rho^2_{ac_1|Z}$ would be larger than $\rho^2_{ac_3|Z}$. This partially justifies the intuitive argument given in Greenland (2003) (see also Greenland and Pearl 2011).

2.2 Comparing conditional dependence with fixed correlates

Here two components a and c of V are held fixed. We consider the variation in $\rho^2_{ac|Z}$ for different subsets Z of V. Depending on the nature of pairwise unconditional association between a, c and the sets conditioned on, three situations may arise.

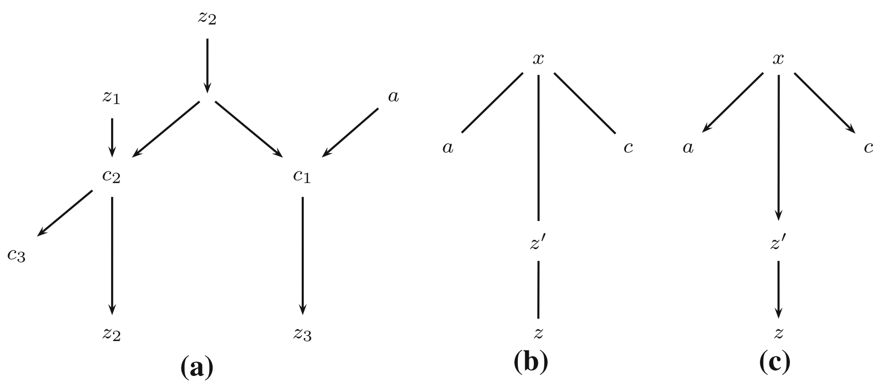


Fig. 1 **a** A polytree, $Z = \{z_1, z_2, z_3, z_4\}$, $\rho^2_{ac_3|Z} \le \rho^2_{ac_2|Z}$, however $\rho^2_{ac_3|Z} \le \rho^2_{ac_1|Z}$ may not hold. **b** an UG satisfying the conditions of Theorem 2, $\rho^2_{ac} \ge \rho^2_{ac|z} \ge \rho^2_{ac|z'}$ and $\rho^2_{ax} \ge \rho^2_{ax|z} \ge \rho^2_{ax|z'}$. Further, from Theorem 1, $\rho^2_{ac} \le \rho^2_{ax}$, $\rho^2_{ac|z} \le \rho^2_{ax|z}$ and $\rho^2_{ac|z'} \le \rho^2_{ax|z'}$. Exactly the same conclusions hold on the DAG in **c**

2.2.1 Situation 1

The components a, c, z and z' are unconditionally pairwise dependent.

Theorem 2 *Suppose for some* $x, a \perp\!\!\!\perp c|x$ *and* $ac \perp\!\!\!\perp z|x$. *Then* $\rho^2_{ac|z} \leq \rho^2_{ac}$. *In addition, if* $ac \perp\!\!\!\perp z'|z$, *then* $\rho^2_{ac|z} \leq \rho^2_{ac|z'} \leq \rho^2_{ac}$.

The conditions of Theorem 2 can be represented by several graphical Markov models, e.g. undirected graphs, directed acyclic graphs etc. The conditional independence conditions imply that a, c and z have to be pairwise separated given x and z' has to be separated from a and c given z.

The first part shows that under these conditions the dependence of a on c always reduces on conditioning. For tree and polytree models the conclusion of the second part can be intuitively explained. Notice that, by assumption $\rho^2_{ac} \geq \rho^2_{ac|x} = 0$ and the separation criteria imply that z' is farther away from x than z. Thus z' has less information about x than z. So $\rho^2_{ac|z'}$ should be closer to ρ^2_{ac} than $\rho^2_{ac|z}$. In other words, conditioning on the vertices farther away from the path between a and c increases the degree of association.

2.2.2 Situation 2

The correlates a and c are independent, but both are dependent on the sets conditioned on.

Theorem 3 *Suppose* $a \perp\!\!\!\perp c$ *and for some* x, *the condition* $ac \perp\!\!\!\perp zB|x$ *holds. Then* $\rho^2_{ac|B} \leq \rho^2_{ac|Bz}$. *Moreover, if* $z' \perp\!\!\!\perp acB|z$ *holds, then* $\rho^2_{ac|B} \leq \rho^2_{ac|Bz'} \leq \rho^2_{ac|Bz}$.

By assumption $0 = \rho^2_{ac} \leq \rho^2_{ac|B}$. Thus the first conclusion implies that conditioning on a larger set implies stronger association. On an UG, the condition $a \perp\!\!\!\perp c$ implies that a and c cannot be connected. Thus UGs are not useful to represent the conditions in Theorem 3. They are satisfied by several other graphical Markov models like DAGs, MAGs etc.

For polytree models (see Fig. 2a) the conclusions of Theorem 3 can be intuitively explained as well. As before, one can conclude z' is farther away from x and therefore has less information about x than z, $\rho^2_{ac|x} \neq 0$ but $\rho^2_{ac} = 0$. Thus by the same argument as for Theorem 2, conditioning on B and z' should produce weaker association than B and z.

In the graph in Fig. 2b the marginal covariance matrix of a, c, x and y satisfy the conditions of Theorem 3. Thus, $\rho^2_{ac|y} \leq \rho^2_{ac|x}$. The graph in Fig. 2c is a mixed ancestral graph (notice the $\leftrightarrow$ edge between y_1 and y_2; Richardson and Spirtes 2002). Here the marginal covariance matrix of a, c, x_2, z and z' would satisfy the conditions of Theorem 3 (see supplement). So we conclude that $\rho^2_{ac|z'} \leq \rho^2_{ac|z} \leq \rho^2_{ac|x_2}$.

2.2.3 Situation 3

At least one of a and c is independent of both the sets conditioned on.

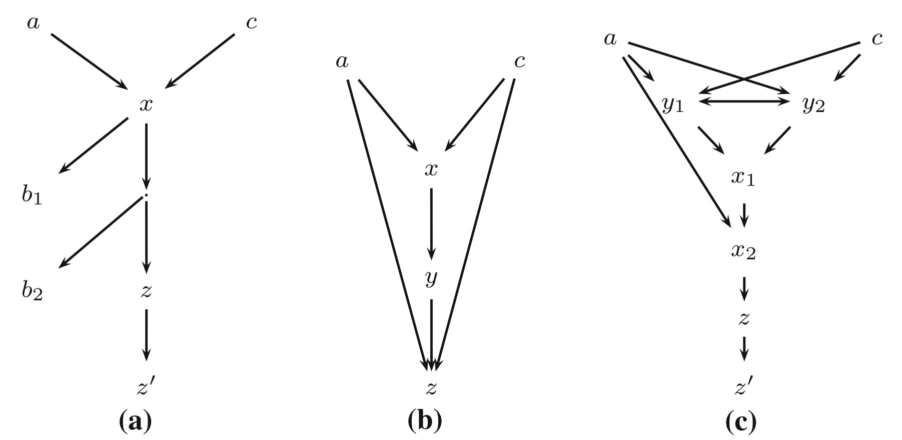


Fig. 2 Graphical models satisfying the conditions of Theorem 3. In each graph $a \perp\!\!\!\perp c$. The graph in **a** is a polytree. Here $B = \{b_1, b_2\}$ and $\rho^2_{ac|B} \le \rho^2_{ac|Bz'} \le \rho^2_{ac|Bz}$ holds. In **b** it follows that $\rho^2_{ac|y} \le \rho^2_{ac|x}$ (cf. Wermuth and Cox 2008). The graph in **c** is a mixed ancestral graph (Richardson and Spirtes 2002) where $\rho^2_{ac|z'} \le \rho^2_{ac|z}$ always holds

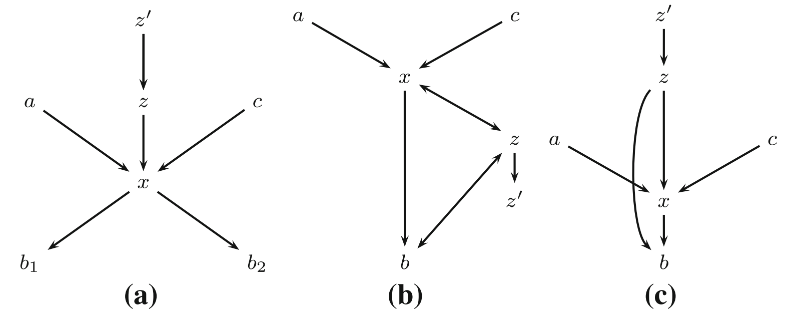


Fig. 3 Graphical models satisfying the conditions of Theorem 4. Each model satisfies the condition (i) of the theorem. **a** is a polytree on which $aczz' \perp\!\!\!\perp \{b_1, b_2\}|x$ holds. In **b**, $ac \perp\!\!\!\perp b|x$, but $ac \not\perp\!\!\!\perp b|zx$. In **c**, $ac \perp\!\!\!\perp b|zx$ but $ac \not\perp\!\!\!\perp b|x$. From Theorem 4 it follows that $\rho^2_{ac|B} \le \rho^2_{ac|Bz'} \le \rho^2_{ac|Bz}$

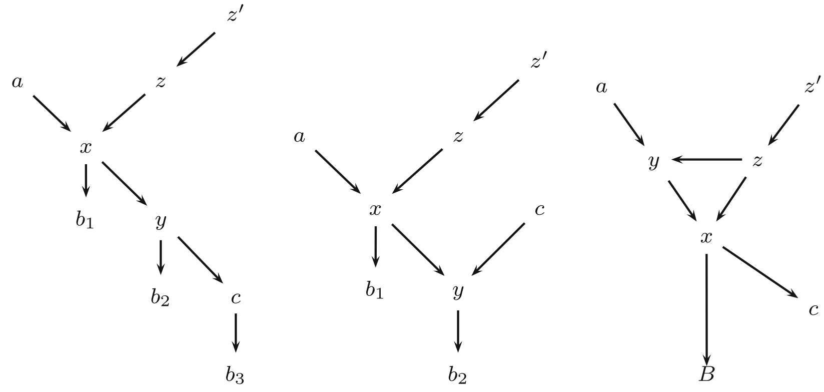
Theorem 4 *Suppose $a \perp\!\!\!\perp z$. Let for some x, Σ satisfies one of the following two* ((i), (ii)) *conditions:*

(i) *$c \perp\!\!\!\perp az$ and one of the following six conditions* (a) $az \perp\!\!\!\perp B|x$, (b) $az \perp\!\!\!\perp B|cx$, (c) $cz \perp\!\!\!\perp B|x$, (d) $cz \perp\!\!\!\perp B|ax$, (e) $ac \perp\!\!\!\perp B|x$ *and* (f) $ac \perp\!\!\!\perp B|xz$ *holds,*

(ii) $az \perp\!\!\!\perp cB|x$.

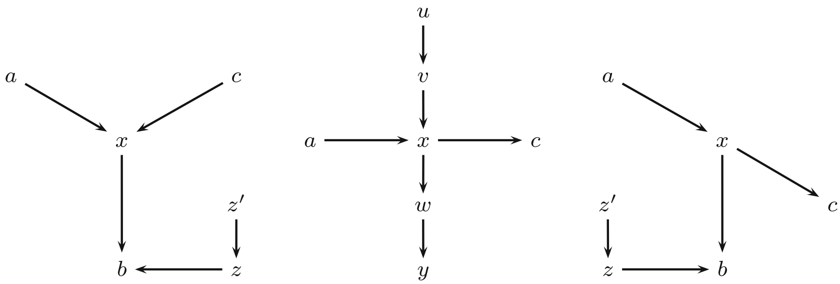
Then $\rho^2_{ac|B} \le \rho^2_{ac|Bz}$. Further, if $z' \perp\!\!\!\perp acB|z$ holds, then in both cases, $\rho^2_{ac|B} \le \rho^2_{ac|Bz'} \le \rho^2_{ac|Bz}$.

The difference between the conditions (i) and (ii) in Theorem 4 is illustrated in Figs. 3a and 4a. Under condition (i), $c \perp\!\!\!\perp z$ but the relation $c \not\perp\!\!\!\perp z|x$ does not necessarily hold. On the other hand, under condition (ii), $c \perp\!\!\!\perp z|x$ but c may not be independent z unconditionally.



(a) **(b)** **(c)**

Fig. 4 Graphical models satisfying the conditions (ii) of Theorem 4. In each graph the condition $az \perp\!\!\!\perp cB|x$ holds. In **b** $az \perp\!\!\!\perp \{b_1, b_2\}|cx$ also holds. The graphs in **a** ($B = \{b_1, b_2, b_3\}$) and **b** ($B = \{b_2, b_3\}$) are polytrees. On each $\rho^2_{ac|B} \leq \rho^2_{ac|Bz'} \leq \rho^2_{ac|Bz}$ hold



(a) **(b)** **(c)**

Fig. 5 **a** A DAG not considered by Chaudhuri and Richardson (2003). From Theorem 4, it follows that $\rho^2_{ac|b} \leq \rho^2_{ac|bz'} \leq \rho^2_{ac|bz}$. **b** A DAG to illustrate the contrast in the conclusion of Theorem 2 and Theorem 4(ii). Here $\rho^2_{ac|v} \geq \rho^2_{ac|u} \geq \rho^2_{ac} \geq \rho^2_{ac|y} \geq \rho^2_{ac|w} \geq \rho^2_{ac|x} = 0$. From Theorem 2, on the DAG in **c** it follows that $\rho^2_{ac|b} \geq \rho^2_{ac|bz'} \geq \rho^2_{ac|bz}$ always hold

The six conditions in (i) are in general distinct. As for example, from m-connection rules (Richardson and Spirtes 2002) the MAG in Fig. 3b we get (note the paths $(a, c) \leftrightarrow x \leftrightarrow z \leftrightarrow b$) $ac \perp\!\!\!\perp b|x$ but $ac \not\!\perp\!\!\!\perp b|zx$ (see supplement). On the other hand on the DAG in Fig. 3c clearly $ac \perp\!\!\!\perp b|zx$ but $ac \not\!\perp\!\!\!\perp b|x$. Similar examples for other four conditions can be drawn.

Theorem 4 goes beyond the DAGs considered by Chaudhuri and Richardson (2003). One example is considered in Fig. 5a. Here $a \perp\!\!\!\perp c$, $ac \perp\!\!\!\perp z$ and both $ac \perp\!\!\!\perp b|x$ and $ac \perp\!\!\!\perp b|xz$ holds. Consequently, from Theorem 4, the relationship $\rho^2_{ac|b} \leq \rho^2_{ac|bz'} \leq \rho^2_{ac|bz}$ follows. Note that z is not an *ancestor* of x but an *ancestor* of b and consequently, $zz' \perp\!\!\!\perp x$ also holds. Chaudhuri and Richardson (2003) explicitly exclude conditioning vertices which are independent of x.

Corollary 1 *If $B = \emptyset$, Under all conditions of Theorem* 4(i), $\rho^2_{ac|z} = \rho^2_{ac|z'} = \rho^2_{ac} = 0$. *Under condition* (ii), $\rho^2_{ac|z} \geq \rho^2_{ac|z'} \geq \rho^2_{ac}$.

2.3 Comparison between Theorems 2 and 4 for polytree models

For polytree models, in view of Theorem 2, the conclusion of Theorem 4(ii) is a bit counterintuitive. Note that, under (ii), $\rho^2_{ac|x} = 0$, which is same as in Theorem 2. However, unlike the latter, conditioning on vertices farther away produce a weaker squared correlation in this case. The difference seems to be that in Theorem 2 $a \not\perp\!\!\!\perp z$, but we assume $a \perp\!\!\!\perp z|x$. In contrast, Theorem 4 assumes that $a \perp\!\!\!\perp z$, but in (ii), the condition $a \perp\!\!\!\perp z|x$ does not hold. As an illustration of this contrast we consider the graph in Fig. 5b. From Theorem 2 and Corollary 1 it follows that the relationship $\rho^2_{ac|v} \geq \rho^2_{ac|u} \geq \rho^2_{ac} \geq \rho^2_{ac|y} \geq \rho^2_{ac|w} \geq \rho^2_{ac|x} = 0$ holds.

Another such example can be constructed from the DAG in Fig. 5a. We have argued above that from Theorem 4 it follows that $\rho^2_{ac|b} \leq \rho^2_{ac|bz'} \leq \rho^2_{ac|bz}$. In the DAG in Fig. 5c the relation $a \perp\!\!\!\perp c$ has been replaced by $a \perp\!\!\!\perp c|x$. From the rules of d-separation $a \perp\!\!\!\perp c|bx, ac \perp\!\!\!\perp z|bx$ and $acb \perp\!\!\!\perp z'|z$ (see Definition 4). Thus after conditioning on b, the Covariance matrix of a, x, c, z and z' satisfies the conditions of Theorem 2. So the qualitative comparison holds, but in contrast to Fig. 5a, it follows that $\rho^2_{ac|b} \geq \rho^2_{ac|bz'} \geq \rho^2_{ac|bz}$.

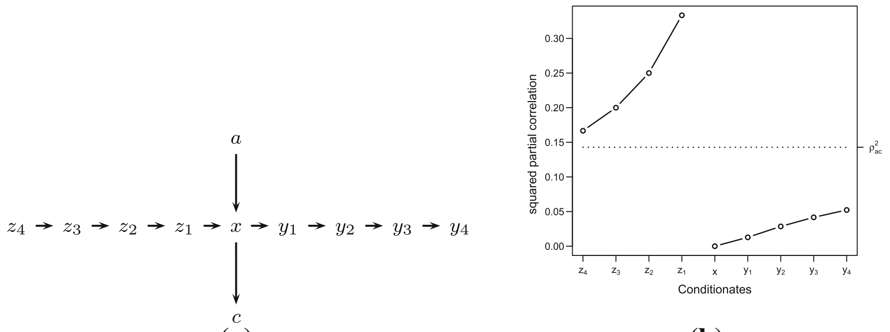
2.4 Comparison between $\rho^2_{ac|x}$ and $\rho^2_{ac|Bz}$

If $z = x$, in Theorem 4 in all case $a \perp\!\!\!\perp Bcz$, so $\rho^2_{ac|z'} = \rho^2_{ac|Bz} = \rho^2_{ac|B} = 0$. When $x \in V \setminus z$, comparison between $\rho^2_{ac|x}$ and $\rho^2_{ac|Bz}$ does not directly follow from Theorem 4. Under condition (ii), $a \perp\!\!\!\perp c|x, 0 = \rho^2_{ac|x} \leq \rho^2_{ac|Bz}$ for any z. However, under the conditions (i), $\rho^2_{ac|x}$ and $\rho^2_{ac|Bz}$ may not be qualitatively compared. We show this fact in the following theorem.

Theorem 5 *Suppose $a \perp\!\!\!\perp z, c \perp\!\!\!\perp az$, and $acz \perp\!\!\!\perp B|x$, then $\rho^2_{ac|Bz} \geq \rho^2_{ac|x}$, iff*

$$\left(\sigma_{xx} + \frac{\sigma^2_{xz}}{\sigma_{zz}}\right) \Sigma_{xB} \Sigma^{-1}_{BB} \Sigma_{Bx} \geq \sigma^2_{xx}, \text{ or equivalently } \frac{\sigma_{xx} - \sigma_{xx|B}}{\sigma_{xx}} \geq \frac{\sigma_{zz|B}}{\sigma_{zz}}.$$

Theorems 2, 3, 4 and 5 have a curious implication on polytree models. Notice that in Theorems 2 and 3 the vertex z is in the set of *descendants* of vertex x (see Figs. 1c, 2a), whereas in Theorem 4, z may be a *parent* of x. The curious fact is that, on a polytree the squared partial correlations given the descendants of x cannot be compared with the squared partial correlations given the parents (or more generally given the *ancestors of the parents* of x). Furthermore, the behaviour of $\rho^2_{ac|x}$ is a continuation of the behaviour of squared partial correlations given its descendants. In other words, on polytrees, conditioning on the vertices "above"



(a) **(b)**

Fig. 6 **a** A polytree and **b** the value of $\rho^2_{ac|i}$ for $i \in \{\emptyset, z_4, z_3, z_2, z_1, x, y_1, y_2, y_3, y_4\}$. Each parameter is fixed at 1. **b** Illustrates the discontinuous drop in $\rho^2_{ac|i}$ as we move from z_1 to x along the z_4 to y_4 path

the path has different nature than conditioning on the vertices "below" or "on" the path.

We present an illustrative example in Fig. 6. We consider the polytree in Fig. 6a. In Fig. 6b we plot the values of $\rho^2_{ac|i}$ for $i \in \{\emptyset, z_4, z_3, z_2, z_1, x, y_1, y_2, y_3, y_4\}$. All parameter values are fixed at 1. As predicted from Theorem 4 the squared partial correlation increases from $i = z_4$ to $i = z_1$ and from Corollary 1 each of them are larger than ρ^2_{ac}. However, From Theorem 3, $\rho^2_{ac|i}$ increases as we move from x to y_4 and each of them are smaller that ρ^2_{ac}. Thus the squared partial correlation drops discontinuously as we move from z_1 to x along the z_4 to y_4 path.

2.5 Further generalisations on comparison with fixed correlates

Suppose $Z_1 = \{z_{11}, z_{12}, \ldots, z_{1n}\}$ and $Z_2 = \{z_{21}, z_{22}, \ldots, z_{2n}\}$ are two conditionates of cardinality n. Then for fixed correlates a and c, one can write:

$$\frac{\rho^2_{ac|Z_1}}{\rho^2_{ac|Z_2}} = \prod_{i=1}^{n} \frac{\rho^2_{ac|z_{21},z_{22},\ldots,z_{2(i-1)},z_{1i},z_{1(i+1)},\ldots,z_{1n}}}{\rho^2_{ac|z_{21},z_{22},\ldots,z_{2(i-1)},z_{2i},z_{1(i+1)},\ldots,z_{1n}}} \tag{2}$$

Clearly $\rho^2_{ac|Z_1} \leq \rho^2_{ac|Z_2}$ holds if each factor in the R.H.S. of (2) is bounded by 1.

Note that in each factor in (2) the conditionate in the numerator and the denominator differ only in one element. Thus in order to qualitatively compare $\rho^2_{ac|Z_1}$ and $\rho^2_{ac|Z_2}$ it is sufficient to find a x_i for each factor such that z_{1i} and z_{2i} satisfy the conditions of one of the Theorems 2–4, possibly with $B \subseteq \{z_{21}, z_{22}, \ldots, z_{2(i-1)}, z_{1(i+1)}, \ldots, z_{1n}\}$ whenever necessary.

Using the factorisation in (2) and Theorems 2–4, structural and path based rules for comparison may be postulated for several graphical models. The choice of x_i and these path based rules depend on the structure of association of the whole vector V. We consider the tree models below.

3 Application to tree models

Let $G = (V, E)$ be a tree with vertex set V and edge set E. For vertices $x \in V$ and $y \in V$, ${}_x\pi_y$ denote the unique path joining x and y, which we define as:

$$\begin{aligned} {}_x\pi_y = \{x = v_1, v_2, \ldots, v_{k-1}, v_k = y \text{ such that there is an edge between}\\ v_i \text{ and } v_{i+1}, \text{ for each } i = 1, 2, \ldots, k-1\}. \end{aligned}$$

Notice that, by the above definition ${}_x\pi_y$ is a subset of V which contains the end points x and y. Since G is a tree, it has only one connected component and therefore any two vertices x and y are connected by an unique ${}_x\pi_y$.

Definition 1 Two vertices a and c on an undirected graph G is said to be *separated* given a subset Z of $V \setminus \{a, c\}$ if each path π between a and c intersects Z. Two subsets A and C of V are separated given $Z \subseteq V \setminus (A \cup C)$ if Z separates each $a \in A$ from each $c \in C$. Two subset A and C of V are connected given a subset Z if they are not separated given Z.

Clearly on a tree a and c are separated given each $x \in {}_a\pi_c \setminus \{a, c\}$. On the other hand since any two vertices a and c are connected by an unique path, a and c cannot be separated given the $\emptyset$.

The *separation criterion* described above associates a set of conditional independence relations with G. This set is described by a collection of *triples*.

$$\mathfrak{I}(G) = \{\langle T_1, T_2 \mid T_3\rangle, \text{ where } T_1 \dot\cup T_2 \dot\cup T_3 \subseteq V \text{ such that } T_1 \perp\!\!\!\perp T_2 | T_3\}. \tag{3}$$

The association of the separation criterion with $\mathfrak{I}(G)$ can be described as follows:

$$\langle T_1, T_2 \mid T_3\rangle \Leftrightarrow T_1 \text{ is separated from } T_2 \text{given } T_3 \text{ in } G.$$

If $V \sim N(0, \Sigma)$, then Σ satisfies all conditional independence relationships in $\mathfrak{I}(G)$. This implies that if $\Lambda = \Sigma^{-1}$, for each $\langle T_1, T_2 \mid T_3\rangle \in \mathfrak{I}(G)$, $\Lambda_{T_1T_2} = 0$.

We now define formal operation of conditioning for independence model $\mathfrak{I}(G)$, on subsets of V.

Definition 2 An independence model $\mathfrak{I}(G)$ *after conditioning on a subset* Z is the set of triples defined as follows:

$$\mathfrak{I}(G)\Big[^{Z} \equiv \Big\{\langle T_1, T_2 \mid T_3\rangle \;\Big|\; \langle T_1, T_2 \mid T_3 \cup Z\rangle \in \mathfrak{I}(G);\; (T_1 \cup T_2 \cup T_3) \cap Z = \emptyset\Big\}. \tag{4}$$

Thus if $\mathfrak{I}(G)$ contains the independence relations satisfied by a $N(0, \Sigma)$ on G, then $\mathfrak{I}(G)\Big[^{Z}$ constitutes the subset of independencies holding among the variables in $Z^c = V \setminus Z$, after conditioning on Z. Let G_{Z^c} be the subgraph of G with vertex set Z^c and edge set consisting of all edges in E between the vertices in Z^c. The following Lemma makes the connection between $\mathfrak{I}(G)\Big[^{Z}$ and $\mathfrak{I}(G_{Z^c})$.

Lemma 1 *Suppose $G = (V, E)$ is a tree. Let a, c be two distinct vertices, $Z \subseteq V\backslash\{a, c\}$ and $Z^c = V\backslash Z$. Then*

$$\Im(G)\left[^{Z} = \Im(G_{Z^c}).\right. \tag{5}$$

Lemma 1 holds for any UG. It implies that the conditioning on Z does not add or delete any edge in G_{Z^c}, so if G is tree $\Im(G)\left[^{Z}\right.$ can be represented by a forest. The inverse of conditional covariance matrix of Z^c given Z is simply $\Lambda_{Z^c Z^c}$.

Separation ensures conditional independence, but if even if the separation fails the corresponding conditional covariance can still be zero (implying conditional independence for Gaussian random variables) because of the parameter values. However, Theorem 2 is still valid in these cases.

For a fixed conditionate the rules for comparing squared partial correlations on trees follows easily from Theorem 1 and the separation criterion.

Theorem 6 *Suppose that, on a Gaussian tree G, the vertices a, c, c' are such that $c \in {}_a\pi_{c'}$. Then for any $Z \subseteq V$, $\rho^2_{ac'|Z} \leq \rho^2_{ac|Z}$.*

For fixed correlates a and c and two sets Z_1 and Z_2 of cardinality more than one, $\rho^2_{ac|Z_1}$ and $\rho^2_{ac|Z_2}$ can be compared qualitatively. The following result describes a sufficient condition.

Theorem 7 *Let $G = (V, E)$ be a Gaussian tree. Suppose a and c are two vertices on G and Z_1 and Z_2 are two subsets of V such that $ac \perp\!\!\!\perp Z_2|Z_1$. Then $\rho^2_{ac|Z_1} \leq \rho^2_{ac|Z_2}$.*

From the separation criterion described above, it follows that the vertices a and c separated from Z_2 given Z_1 implies $ac \perp\!\!\!\perp Z_2|Z_1$ and therefore $\rho^2_{ac|Z_1} \leq \rho^2_{ac|Z_2}$. The following Corollary gives the corresponding sufficient condition in terms of paths:

Corollary 2 *Suppose Z_1 and Z_2 are two subsets of V, such that for each vertex $z_2 \in Z_2$, both the paths ${}_a\pi_{z_2}$ and ${}_c\pi_{z_2}$ intersect Z_1, then $\rho^2_{ac|Z_1} \leq \rho^2_{ac|Z_2}$.*

Notice that, Theorem 7 is more general than Corollary 2, the Theorem covers the cases when the conditional independence holds due to the choices of parameters as well. The result in Theorem 7 is also complete in the following sense.

Theorem 8 *Suppose $G = (V, E)$ is a Gaussian tree. Let $Z_1, Z_2 \subseteq V$ such that $ac \not\!\perp\!\!\!\perp Z_2|Z_1$ and $ac \not\!\perp\!\!\!\perp Z_1|Z_2$. Further, suppose that $(Z_1 \cup Z_2) \cap {}_a\pi_c = \emptyset$. Then there exists Σ_1 such that $\rho^2_{ac|Z_1} > \rho^2_{ac|Z_2}$ and Σ_2 such that $\rho^2_{ac|Z_2} > \rho^2_{ac|Z_1}$.*

Finally, Theorem 6 and the Corollary 2 can be combined to a general rule for comparing squared partial correlation on trees.

Corollary 3 *Suppose a, c, c' are three vertices on a Gaussian tree G and Z, Z' are two subsets of the vertex set V. Further, assume that $c \in {}_a\pi_{c'}$ and the vertices a and c' are separated from Z given Z'. Then $\rho^2_{ac'|Z'} \leq \rho^2_{ac|Z}$.*

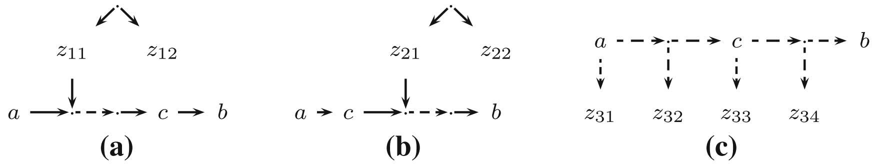


Fig. 7 Examples of polytrees satisfying the conditions of Theorem 9 below. In each, $a \in \mathrm{an}(c)$ and $c \in \mathrm{an}(b)$. In **a** z_{11} and z_{12} satisfy condition 1 and $\rho^2_{ac|b} < \rho^2_{ac|bz_{12}} < \rho^2_{ac|bz_{11}}$ (from Theorem 4(ii)). In **b** z_{21} and z_{22} satisfy condition 2. So $\rho^2_{ac|bz_{21}} < \rho^2_{ac|bz_{22}} < \rho^2_{ac|b}$ (see Theorem 2 and Fig. 5c). Each $z_{3k}, k = 1, \dots, 4$, in **c** satisfy condition 2, i.e. $\rho^2_{ac|z_{3k}} < \rho^2_{ac|b}$. Note that, b cannot be in $\mathrm{an}(z)$, otherwise $ac \perp\!\!\!\perp z|b$ and $\rho^2_{ac|b} = \rho^2_{ac|bz}$

4 Application to polytree models and model selection

A polytree is a DAG such that if we substitute all its directed edges with undirected ones, the resulting graph (i.e. its skeleton) would be a tree. Thus on a polytree two vertices x and y can have at most one path ${}_x\pi_y$ connecting them. Here, on a connecting path we disregard the direction of the individual edges.

A vertex y is an ancestor of a vertex x, if either $y = x$ or x can be reached from y by following the arrowheads of a directed path (i.e. the path $y \to v_1 \to v_2 \dashrightarrow v_k \to x$ exits). The collection of all ancestors of x is denoted by $\mathrm{an}(x)$. Furthermore, for a set of vertices X we define $\mathrm{an}(X) = \cup_{x \in X} \mathrm{an}(x)$.

Theorem 9 *Suppose that on a Gaussian polytree $a \neq c \neq b, a \in \mathrm{an}(c)$ and $c \in \mathrm{an}(b)$. Further let, for some vertex z, $\rho^2_{ac|bz} \neq \rho^2_{ac|b}$. Then*

1. $\rho^2_{ac|bz} > \rho^2_{ac|b}$, *iff $a \perp\!\!\!\perp z$ and $c \not\perp\!\!\!\perp z$.*
2. $\rho^2_{ac|bz} < \rho^2_{ac|b}$ *iff either $c \perp\!\!\!\perp z$ or $a \not\perp\!\!\!\perp z$.*

The condition $\rho^2_{ac|bz} \neq \rho^2_{ac|b}$ is required in Theorem 9. This implies $ac \not\perp\!\!\!\perp z \mid b$. So $b \notin \mathrm{an}(z)$. It can further be shown (see the proof) that the polytree structure implies $ac \perp\!\!\!\perp z$ iff $c \perp\!\!\!\perp z$. Thus the right hand side of condition 2 above equivalently means that either both a and c are independent of z or none of them are independent of z. Examples of graphs satisfying the conditions 1 and 2 can be found in Fig. 7.

Theorem 9 has applications in model selection. An example occurs in the mapping of river flow networks. Figure 8 (Jarvie et al. 2005) presents a schematic diagram of the network of the Avon basin in Hampshire, England. Suppose that it is known that none of the rivers involved have a distributary. Clearly the network, with the direction of the water flow forms a polytree. Measurements can be taken at points a (Netheravon), b (Christchurch), c (Amesbury), d (Downstream of Salisbury STW), e (Longford) and z (Chitterne). However, because of practical considerations we suppose that the measurements are taken when water at Christchurch (b) touches certain levels. Lets assume $\rho^2_{ax|b} \neq \rho^2_{ax|bz}$ for $x = c, d, e$. We want to know where does the stream from z, i.e. Chitterne meets river Avon.

It is clear that since the observations are all conditional on the water level at b, in the data neither $z \not\perp\!\!\!\perp a$ nor $z \not\perp\!\!\!\perp c$. However, from Theorem 2 and Theorem 4,

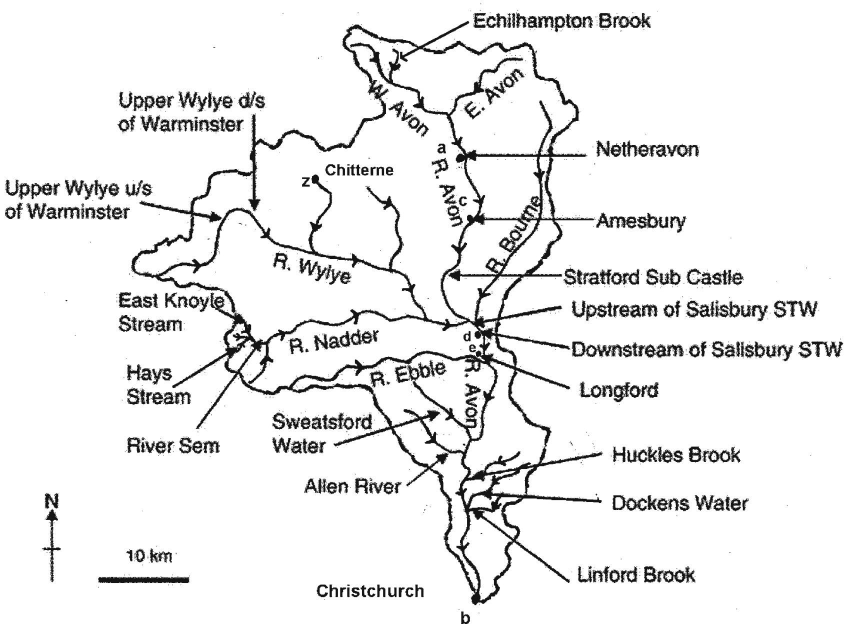


Fig. 8 An illustration of the results in Theorem 9 on the river network of Avon river, Hampshire, England (obtained from Jarvie et al. 2005)

see also Fig. 5c, it follows that $\rho^2_{ac|bz} < \rho^2_{ac|b}, \rho^2_{ad|bz} > \rho^2_{ad|b}$ and $\rho^2_{ae|bz} > \rho^2_{ae|b}$. From condition 2 of Theorem 9 it follows that either both a and c are independent of z or none of them are. On the other hand, condition 1. implies that $a \perp\!\!\!\perp z$ but d and e are not independent of z. If none of a and c are independent of z, the point z must be on a distributary stream or on a tributary which meets Avon north of a (Netheravon). However, by assumption there is no distributary stream. Furthermore, if the tributary from z meets Avon somewhere north of a, by Theorem 2 both $\rho^2_{ad|bz} < \rho^2_{ad|b}$ and $\rho^2_{ae|bz} < \rho^2_{ae|b}$ must hold. This is a contradiction. Thus $ac \perp\!\!\!\perp z$ must hold. So from Theorem 9 we see that the stream from Chitterne i.e. z meets Avon somewhere between Amesbury i.e. c and Downstream of Salisbury STW i.e. d.

5 Discussion

Qualitative comparison may be possible under other sets of conditional independence relations. However, fairly simple examples (Chaudhuri 2013) show that violation of many of these conditions in the sets described above lead to non-qualitative comparison of mutual informations. The requirement of a single component x cannot be relaxed.

The results in Sect. 2 are sufficient for postulating path based rules for comparison on polytree models as well. Since the edges on a polytree are directed, these rules are more involved than those for trees (Chaudhuri and Richardson 2003).

Comparison of mutual information with a fixed conditionate holds for any distribution. In fact, the results with fixed correlates are based on the positive-definiteness of the covariance matrix and extend to non-Gaussian distributions as well. However, inequalities for squared partial correlation would not translate to mutual information for such random variables. These results may be applicable to causal model selections among non-Gaussian variables (e.g. Shimizu et al. 2006).

It can be shown that, although the comparisons with a fixed conditionate do not hold, but absolute values of partial regression coefficients can be qualitatively compared for fixed correlates under the same conditions (Chaudhuri and Tan 2010).

Rules for signed comparisons of partial correlation and regression coefficients can be developed from these results. Such results might be useful in identifying hidden variables in Factor models (Bekker and Leeuw 1987; Drton et al. 2007; Xu and Pearl 1989; Spirtes et al. 2000) and in recovering population covariance matrix for one-factor models in presence of selection bias (Kuroki and Cai 2006).

Appendix: Proofs

Notation For two real numbers a and b, $a \propto^{+} b$ implies that, $\exists\, M > 0$ such that $a = M \cdot b$.

Proposition 1 *Suppose U, V, W are univariate components of a Gaussian random vector with mean μ and positive definite covariance Σ. Assume that $U \perp\!\!\!\perp V|W$. Then $\sigma_{UV} = \sigma_{UW}\sigma_{WV}/\sigma_{WW}$ and $\sigma_{UU} = \sigma^2_{UW}\sigma_{WW}/\sigma^2_{WW} + E[\mathrm{Var}(U|W)]$.*

Proof Trivial.

Suppose K and K' are constants and for some $a, c, d \in V$ and $B \subseteq V \setminus \{a, c, d\}$ (where B may be empty) we denote $M_1 = \sigma_{cd|B}\{\sigma_{ad|B}\sigma_{cc|B} - \sigma_{ac|B}\sigma_{cd|B}\}$, $M_2 = \sigma_{ad|B}\{\sigma_{cd|B}\sigma_{aa|B} - \sigma_{ac|B}\sigma_{ad|B}\}$, $M_3(\alpha) = [(\alpha - K')\sigma_{ac|B}\sigma_{dd|B} - K \cdot \sigma_{ad|B}\sigma_{cd|B}]$ and

$$L(\alpha) = \frac{\{(\alpha - K')\rho_{ac|B} - K\rho_{ad|B}\rho_{cd|B}\}^2}{[\{(\alpha - K') - K\rho^2_{ad|B}\}\{(\alpha - K') - K\rho^2_{cd|B}\}]}. \tag{6}$$

Lemma 2 *Suppose $K > 0$ and for some K' and α, $(\alpha - K') - K\rho^2_{ad|B} > 0$ and $(\alpha - K') - K\rho^2_{cd|B} > 0$.*

Then if $M_1 \cdot M_2 \geq 0$:

1. $\frac{\partial L(\alpha)}{\partial \alpha} = 0$ *if both $M_1 \cdot M_3(\alpha)$ and $M_2 \cdot M_3(\alpha)$ are* 0.
2. $\frac{\partial L(\alpha)}{\partial \alpha}$ *has the same sign as either $M_1 \cdot M_3(\alpha)$ or $M_2 \cdot M_3(\alpha)$, whichever is non-zero.*

Proof Since the denominator of (6) is positive then the sign of $\partial L(\alpha)/\partial\alpha$ is the sign of the numerator of $\partial L(\alpha)/\partial\alpha$. From quotient rule of differentiation and some algebraic manipulation we get:

$$\frac{\partial L(\alpha)}{\partial\alpha} \propto^{+} K[(\alpha - K')\rho_{ac|B} - K\rho_{ad|B}\rho_{cd|B}]$$
$$\times\Big\{[(\alpha - K') - K\rho^2_{ad|B}]\rho_{cd|B}[\rho_{ad|B} - \rho_{ac|B}\rho_{cd|B}]$$
$$+ [(\alpha - K') - K\rho^2_{cd|B}]\rho_{ad|B}[\rho_{cd|B} - \rho_{ac|B}\rho_{ad|B}]\Big\}. \quad (7)$$

Note that $\rho_{cd|B}[\rho_{ad|B} - \rho_{ac|B}\rho_{cd|B}] \propto^{+} M_1$, $\rho_{ad|B}[\rho_{cd|B} - \rho_{ac|B}\rho_{ad|B}] \propto^{+} M_2$ and $[(\alpha - K')\rho_{ac|B} - K\rho_{ad|B}\rho_{cd|B}] \propto^{+} M_3(\alpha)$. By substituting these expressions in (7) and the positivity K, $[(\alpha - K') - K\rho^2_{ad|B}]$ and $[(\alpha - K') - K\rho^2_{cd|B}]$ the result follows. □

Proof of Theorem 1 From the assumption $\mathrm{Inf}(a \perp\!\!\!\perp c'|cZ) = 0$. The rest follows from the identity $\mathrm{Inf}(a \perp\!\!\!\perp c'|cZ) + \mathrm{Inf}(a \perp\!\!\!\perp c|Z) = \mathrm{Inf}(a \perp\!\!\!\perp c|c'Z) + \mathrm{Inf}(a \perp\!\!\!\perp c'|Z)$.[1] □

Note that, from Lněnička and Matúš (2007), assumptions on conditional independence and the conditional correlations do not change if we replace Σ by $J\Sigma J$, where J is the diagonal matrix with $1/\sqrt{\sigma_{vv}}$, $v \in V$. Thus, unless otherwise stated, w.l.g we can assume that the diagonal elements of Σ are all equal to 1 and all the off diagonals are in $(-1, 1)$. That is Σ is the correlation matrix of V, but with an abuse of notation in what follows below, we still denote the correlation of a and c by σ_{ac}.

Proof of Theorem 2 Note that by assumption $\sigma_{ac} = \sigma_{ax}\sigma_{cx}$, $\sigma_{az} = \sigma_{ax}\sigma_{xz}$, $\sigma_{cz} = \sigma_{cx}\sigma_{xz}$, $\sigma_{az'} = \sigma_{ax}\sigma_{xz}\sigma_{zz'}$ and $\sigma_{cz'} = \sigma_{cx}\sigma_{xz}\sigma_{zz'}$.

Part 1. $\rho^2_{ac|z} = \sigma^2_{ac}(1 - \sigma^2_{xz})^2/[(1 - \sigma^2_{ax}\sigma^2_{xz})(1 - \sigma^2_{cx}\sigma^2_{xz})] \le \rho^2_{ac}$.

Part 2. Assume that $x \ne z'$ and consider three non trivial cases as $x = a$, $x = c$ and $x \notin \{a, c\}$. Initially assume that $\sigma_{zz'} \ne 0$. Since $ac \perp\!\!\!\perp z'|z$, using Proposition 1 and the positive definiteness of the covariance matrix together with $\tau^2_z = (1 - \sigma^2_{zz'}) > 0$ and by denoting $\alpha = 1 + (\tau^2_z/\sigma^2_{zz'}) > 1$, with $B = \emptyset$, $K' = 0$, $K = 1$ it follows that $\rho^2_{ac|z'} = L(\alpha)$ for $\alpha \ge 1$ and $\rho^2_{ac|z} = L(1)$. Thus in Lemma 2 using Cauchy Schwartz inequality and $\alpha \ge 1$ it follows that for $x = a$, $M_1 \propto^{+} \sigma_{cx}$, $M_2 = 0$ and $M_3(\alpha) \propto^{+} \sigma_{cx}$, for $x = c$, $M_1 = 0$, $M_2 \propto^{+} \sigma_{ax}$ and $M_3(\alpha) \propto^{+} \sigma_{ax}$ and for $x \notin \{a, c\}$, $M_1 \propto^{+} \sigma_{ax}\sigma_{cx}$, $M_2 \propto^{+} \sigma_{ax}\sigma_{cx}$ and $M_3(\alpha) \propto^{+} \sigma_{cx}\sigma_{ax}$. Thus for all cases $\partial L/\partial\alpha \ge 0$ and the result follows. If $\sigma_{zz'} = 0$, $z \perp\!\!\!\perp z'$ and $z' \perp\!\!\!\perp acz$. Thus $\rho^2_{ac|z'} = \rho^2_{ac}$. The rest follows from part 1. The second inequality follows from part 1 as well. □

Proof of Theorem 3 By assumption $zB \perp\!\!\!\perp ac|x$ and $a \perp\!\!\!\perp c$.

Part 1. It is enough to show that $\sigma^2_{ac|Bz} \ge \sigma^2_{ac|B}$. Using the above relations in Proposition 1 and by denoting $Q_1 = \Sigma_{xB}\Sigma^{-1}_{BB}\Sigma_{Bx}$ and $Q_2 = (\Sigma_{xB}, \sigma_{xz})\Sigma^{-1}_{(Bz)(Bz)}$ $(\Sigma_{xB}, \sigma_{xz})^{\mathrm{T}}$ one gets $\sigma_{ac|Bz} = -\sigma_{ax}\sigma_{cx}Q_2$ and $\sigma_{ac|B} = -\sigma_{ax}\sigma_{cx}Q_1$. Now the proof follows by noting that, $\sigma_{aa} - \sigma^2_{ax}Q_1 = \sigma_{aa|B} \ge \sigma_{aa|Bz} = \sigma_{aa} - \sigma^2_{ax}Q_2$ implies $Q_2 \ge Q_1$.

[1] The author would like to thank the referee for drawing his attention to this equality which improved the original proof immensely.

Part 2. We initially assume that $\sigma_{zz'} \neq 0$. By defining $\tau^2_{z'} = (1 - \sigma^2_{zz'}) > 0, \alpha = (1+(\tau^2_{z'}/\sigma^2_{zz'}))$, $K' = \Sigma_{zB}\Sigma^{-1}_{BB}\Sigma_{Bz} > 0$, $K = (1-K') > 0$ and from the assumption that $z' \perp\!\!\!\perp acB|z$ it follows that $\rho^2_{ac|Bz'} = L(\alpha)$ with $\alpha \geq 1$ and $\rho^2_{ac|Bz} = L(1)$. Further using $ac \perp\!\!\!\perp zB|x$ one can show that $M_1 \propto^+ \sigma_{cx}\sigma_{ax}\sigma^2_{xz|B}$, $M_2 \propto^+ \sigma_{cx}\sigma_{ax}\sigma^2_{xz|B}$ and $M_3(\alpha) \propto^+ -\sigma_{cx}\sigma_{ax}$. Thus from Lemma 2 it follows that $\partial L/\partial\alpha \leq 0$. If $\sigma_{zz'} = 0$, as before $z \perp\!\!\!\perp z'$ and $z' \perp\!\!\!\perp acB$. Thus $\rho^2_{ac|Bz} = \rho^2_{ac|B}$. The result follows from part 1. The first inequality follows from the assumptions and Part 1. □

Proof of Theorem 4 W.l.g. it is enough assume that $x \notin B$. Furthermore, note that $\sigma_{aa|B} \geq \sigma_{aa|Bz}$ and $\sigma_{cc|B} \geq \sigma_{cc|Bz}$, thus for part 1 it is enough to show that under the assumptions $\sigma_{ac|Bz} = m \cdot \sigma_{ac|B}$ for some $m > 1$.

Part 1. Assume that, $a \perp\!\!\!\perp z$ and let (ii) hold, i.e. $cB \perp\!\!\!\perp az|x$. Using Proposition 1 it follows that

$$\begin{aligned}\sigma_{ac|Bz} &= \sigma_{ac|B} + \frac{(\Sigma_{aB}\Sigma^{-1}_{BB}\Sigma_{Bz})(\sigma_{cz} - \Sigma_{cB}\Sigma^{-1}_{BB}\Sigma_{Bz})}{\sigma_{zz|B}}\\ &= \sigma_{ac|B} + \frac{\sigma_{ax}\sigma^2_{xz}Q_1(\sigma_{cx} - \sigma_{cx}Q_1)}{\sigma_{zz|B}}\\ &= \sigma_{ac|B} + \frac{\sigma^2_{xz}Q_1(\sigma_{cx}\sigma_{ax} - \sigma_{cx}\sigma_{ax}Q_1)}{\sigma_{zz|B}} = \sigma_{ac|B}\left(1 + \sigma^2_{zx}Q_1\sigma^{-1}_{zz|B}\right).\end{aligned}$$

Thus $\rho^2_{ac|B} \leq \rho^2_{ac|Bz}$. Under (i) if $c \perp\!\!\!\perp az$, $\sigma_{ac} = \sigma_{zc} = 0$, $\sigma_{ac|B} = -\Sigma_{aB}\Sigma^{-1}_{BB}\Sigma_{Bc}$ and $\sigma_{cz|B} = -\Sigma_{cB}\Sigma^{-1}_{BB}\Sigma_{Bz}$. Now if (i)(a) i.e. $az \perp\!\!\!\perp B|x$ holds:

$$\begin{aligned}\sigma_{ac|Bz} &= \sigma_{ac|B} - \frac{(\Sigma_{aB}\Sigma^{-1}_{BB}\Sigma_{Bz})(\Sigma_{cB}\Sigma^{-1}_{BB}\Sigma_{Bz})}{\sigma_{zz|B}}\\ &= \sigma_{ac|B} - \frac{(\sigma_{ax}\sigma_{xz}Q_1)(\Sigma_{cB}\Sigma^{-1}_{BB}\Sigma_{Bx}\sigma_{xz})}{\sigma_{zz|B}} = \sigma_{ac|B}\left(1 + \sigma^2_{zx}Q_1\sigma^{-1}_{zz|B}\right).\end{aligned} \tag{8}$$

Under (i)(b) i.e. $az \perp\!\!\!\perp B|cx$ notice that from Proposition 1:

$$\begin{aligned}\Sigma_{aB} &= \Sigma_{a(xc)}\Sigma^{-1}_{(xc)(xc)}\Sigma_{(xc)B} = [\sigma_{ax}, 0]\Sigma^{-1}_{(xc)(xc)}\Sigma_{(xc)B}\\ &= \sigma_{ax}[1, 0]\Sigma^{-1}_{(xc)(xc)}\Sigma_{(xc)B} = \sigma_{ax}\mathcal{Q}_{cxB}.\end{aligned}$$

Here $\mathcal{Q}_{cxB} = [1, 0]\Sigma^{-1}_{(xc)(xc)}\Sigma_{(xc)B}$. Similarly it can be shown that, $\Sigma_{zB} = \sigma_{zx}\mathcal{Q}_{cxB}$ and $\sigma_{ac|B} = -\sigma_{ax}\mathcal{Q}_{cxB}\Sigma^{-1}_{BB}\Sigma_{Bc}$. Now by substitution in (8) above we get:

$$\begin{aligned}\sigma_{ac|Bz} &= \sigma_{ac|B} - \frac{\sigma_{ax}(\mathcal{Q}_{cxB}\Sigma^{-1}_{BB}\mathcal{Q}^{\mathrm{T}}_{cxB})(\Sigma_{cB}\Sigma^{-1}_{BB}\mathcal{Q}^{\mathrm{T}}_{cxB})\sigma^2_{zx}}{\sigma_{zz|B}}\\ &= \sigma_{ac|B} - \frac{(\sigma_{zx}\mathcal{Q}_{cxB}\Sigma^{-1}_{BB}\mathcal{Q}^{\mathrm{T}}_{cxB}\sigma_{zx})(\Sigma_{cB}\Sigma^{-1}_{BB}\mathcal{Q}^{\mathrm{T}}_{cxB}\sigma_{ax})}{\sigma_{zz|B}}\end{aligned}$$

$$= \sigma_{ac|B} + \frac{(\Sigma_{zB}\Sigma_{BB}^{-1}\Sigma_{Bz})\sigma_{ac|B}}{\sigma_{zz|B}} = \sigma_{ac|B}\left\{1 + \left(\Sigma_{zB}\Sigma_{BB}^{-1}\Sigma_{Bz}\right)\sigma_{zz|B}^{-1}\right\}.$$

The proofs for (i)(c) and (i)(d) are similar.

If (i)(e) i.e. $ac \perp\!\!\!\perp B|x$ holds, $\sigma_{ac|B} = -\sigma_{ax}\sigma_{cx}Q_1$ and using Proposition 1 we get,

$$\begin{aligned}\sigma_{ac|Bz} = \sigma_{ac|B} - \frac{(-\Sigma_{aB}\Sigma_{BB}^{-1}\Sigma_{Bz})(-\Sigma_{cB}\Sigma_{BB}^{-1}\Sigma_{Bz})}{\sigma_{zz|B}} \\ = \sigma_{ac|B} - \sigma_{ax}\sigma_{cx}(\Sigma_{xB}\Sigma_{BB}^{-1}\Sigma_{Bz})^2\sigma_{zz|B}^{-1} \\ = \sigma_{ac|B}\left\{1 + (\Sigma_{xB}\Sigma_{BB}^{-1}\Sigma_{Bz})^2/(Q_1\sigma_{zz|B}^{-1})\right\}.\end{aligned}$$

Under condition (i)(f) notice that, $\Sigma_{aB} = \Sigma_{a(xz)}\Sigma_{(xz)(xz)}^{-1}\Sigma_{(xz)B} = \sigma_{ax}[1, 0]\Sigma_{(xz)(xz)}^{-1}$ $\Sigma_{(xz)B} = \sigma_{ax}\mathcal{Q}_{xzB}$. Similarly, $\Sigma_{cB} = \sigma_{cx}\mathcal{Q}_{xzB}$. Now from (8) it follows that:

$$\sigma_{ac|Bz} = \sigma_{ac|B} - \frac{\sigma_{ax}\sigma_{cx}(\mathcal{Q}_{xzB}\Sigma_{BB}^{-1}\Sigma_{Bz})^2}{\sigma_{zz|B}}.$$

Clearly if at least one of $\sigma_{ax}, \sigma_{cx}, \mathcal{Q}_{xzB}$ is zero, the results is trivial. Now suppose none of them equal zero. Then $\mathcal{Q}_{xzB}\Sigma_{BB}^{-1}\mathcal{Q}_{xzB}^{\mathrm{T}} > 0$. Further $\sigma_{ac|B} = -\sigma_{ax}\sigma_{cx}(\mathcal{Q}_{xzB}\Sigma_{BB}^{-1}\mathcal{Q}_{xzB}^{\mathrm{T}})$, which yields

$$\sigma_{ac|Bz} = \sigma_{ac|B}\left\{1 + \frac{(\mathcal{Q}_{xzB}^{T}\Sigma_{BB}^{-1}\Sigma_{Bz})^2}{(\mathcal{Q}_{xzB}\Sigma_{BB}^{-1}\mathcal{Q}_{xzB}^{\mathrm{T}})\sigma_{zz|B}}\right\}.$$

Part 2. Suppose $\sigma_{z'z}^2 > 0$. Let $\tau_{z'}^2 = (1 - \sigma_{z'z}^2) > 0$, $K' = \Sigma_{zB}\Sigma_{BB}^{-1}\Sigma_{Bz}$, $K = (1 - K') > 0$ and $\alpha = 1/\sigma_{z'z}^2 = (1 + \tau_{z'}^2/\sigma_{z'z}^2) \geq 1$. Then from $acB \perp\!\!\!\perp z'|z, a \perp\!\!\!\perp zz'$ it follows that for both cases $\rho_{ac|Bz'}^2 = L(\alpha)$ with $\alpha \geq 1$ and $\rho_{ac|Bz}^2 = L(1)$. Now we consider the four cases in the statement. By denoting $Q_{cx} = \Sigma_{cB}\Sigma_{BB}^{-1}\Sigma_{Bx}$, $Q_{ax} = \Sigma_{aB}\Sigma_{BB}^{-1}\Sigma_{Bx}$ and $\mathcal{Q}_{axB} = [1, 0]\Sigma_{(xa)(xa)}^{-1}\Sigma_{(xa)B}$ it follows that:

$$M_1 \propto^{+} M_2 \propto^{+} \begin{cases} \sigma_{ax}Q_{cx} & \text{if (i), (a)} \\ \sigma_{ax}\mathcal{Q}_{cxB}\Sigma_{BB}^{-1}\Sigma_{Bc} & \text{if (i), (b)} \\ \sigma_{cx}Q_{ax} & \text{if (i), (c)} \\ \sigma_{cx}\mathcal{Q}_{axB}\Sigma_{BB}^{-1}\Sigma_{Ba} & \text{if (i), (d)} \\ \sigma_{ax}\sigma_{cx} & \text{if (i), (e)} \\ \sigma_{ax}\sigma_{cx} & \text{if (i), (f)} \\ -\sigma_{ax}\sigma_{cx} & \text{if (ii)} \end{cases}, \quad M_3(\alpha) \propto^{+} \begin{cases} -\sigma_{ax}Q_{cx} & \text{if (i), (a)} \\ -\sigma_{ax}\mathcal{Q}_{cxB}\Sigma_{BB}^{-1}\Sigma_{Bc} & \text{if (i), (b)} \\ -\sigma_{cx}Q_{ax} & \text{if (i), (c)} \\ -\sigma_{cx}\mathcal{Q}_{axB}\Sigma_{BB}^{-1}\Sigma_{Ba} & \text{if (i), (d)} \\ -\sigma_{ax}\sigma_{cx} & \text{if (i), (e)} \\ -\sigma_{ax}\sigma_{cx} & \text{if (i), (f)} \\ \sigma_{ax}\sigma_{cx} & \text{if (ii)} \end{cases}.$$

Thus from Lemma 2, in all cases $\partial L/\partial\alpha \leq 0$, which completes the proof.

If $\sigma_{zz'} = 0$, then for all cases $\rho^2_{ac|Bz'} = \rho^2_{ac|B}$ and the result follows from Part 1 as before. □

Proof of Corollary 1 If $B = \emptyset$, under (i) from the assumed independence of a, c and z, we get $\sigma_{ac} = \sigma_{az} = \sigma_{cz} = 0$. The result follows from this. Under (ii), $\sigma_{cz} \neq 0$ and from Theorem 4 the result follows. □

Proof of Theorem 5 In this proof we take Σ to be the covariance matrix and not the correlation matrix as above. Using condition $B \perp\!\!\!\perp acz|x$, denoting $\sigma^2_{xx} Q_4 = \Sigma_{xB}\Sigma^{-1}_{BB}\Sigma_{Bx}$, $T = \sigma_{zz}/(\sigma_{zz} - \sigma^2_{xz} Q_4)$ $(T > 0)$ and from Proposition 1 and some simplification we get

$$\frac{\rho^2_{ac|Bz}}{\rho^2_{ac|x}} = \frac{(\sigma_{aa}\sigma_{xx} Q_4 T - \sigma^2_{ax} Q_4 T)(\sigma_{cc}\sigma_{xx} Q_4 T - \sigma^2_{cx} Q_4 T)}{(\sigma_{aa} - \sigma^2_{ax} Q_4 T)(\sigma_{cc} - \sigma^2_{cx} Q_4 T)}.$$

Thus $\rho^2_{ac|Bz} \geq \rho^2_{ac|x}$ iff $\sigma_{xx} Q_4 T \geq 1$ iff $(\sigma_{xx} + \sigma^2_{xz}/\sigma_{zz}) Q_4 \geq 1$. The equivalent expression follows as:

$$\begin{aligned}\frac{\sigma^2_{xz}}{\sigma_{zz}\sigma^2_{xx}}\Sigma_{xB}\Sigma^{-1}_{BB}\Sigma_{Bx} \geq \frac{\sigma_{xx|B}}{\sigma_{xx}} &\Leftrightarrow \frac{1}{\sigma_{zz}}\Sigma_{zB}\Sigma^{-1}_{BB}\Sigma_{Bz} \\ \geq \frac{\sigma_{xx|B}}{\sigma_{xx}} &\Leftrightarrow \frac{\sigma_{xx} - \sigma_{xx|B}}{\sigma_{xx}} \geq \frac{\sigma_{zz|B}}{\sigma_{zz}}.\end{aligned}$$

□

Proof of Lemma 1 We need to show that if T_1, T_2 and T_3, are disjoint subsets of Z^c, then T_1 is connected to T_2 given T_3 in G_{Z^c} iff T_1 is connected to T_2 given $T_3 \cup Z$ in G.

($\Rightarrow$) Suppose T_1 is connected to T_2 given T_3 in G_{Z^c}. So there are $t_1 \in T_1$ and $t_2 \in T_2$ and the path ${}_{t_1}\pi_{t_2}$ such that $\pi \cap T_3 = \emptyset$. Clearly ${}_{t_1}\pi_{t_2}$ is in G and ${}_{t_1}\pi_{t_2} \cap Z = \emptyset$. So ${}_{t_1}\pi_{t_2} \cap \{T_3 \cup Z\} = \emptyset$. This shows T_1 is connected to T_2 given $T_3 \cup Z$ in G.

($\Leftarrow$) Suppose T_1 is connected to T_2 given $T_3 \cup Z$ in G. So there is $t_1 \in T_1$ and $t_2 \in T_2$ and the path ${}_{t_1}\pi_{t_2}$, such that ${}_{t_1}\pi_{t_2} \cap \{T_3 \cup Z\} = \emptyset$. So ${}_{t_1}\pi_{t_2} \cap Z = \emptyset$ and ${}_{t_1}\pi_{t_2} \subseteq Z^c$. Clearly in G_{Z^c}, ${}_{t_1}\pi_{t_2} \cap T_3 = \emptyset$. This shows T_1 is connected to T_2 given T_3 in G_{Z^c}. □

Proof of Theorem 6 From the structure of G and since $c \in {}_a\pi_{c'}$, it easily follows that c' is separated from a given c and Z. The result follows from Theorem 1. □

Proof of Theorem 7 For notational convenience we express the squared partial correlations as functions of the covariance matrix Σ. We need to show that $\rho^2_{ac|Z_1}(\Sigma) \leq \rho^2_{ac|Z_2}(\Sigma)$. W.l.g. we assume that for $i = 1, 2$ there is no $z_i \in Z_i$ such that $ac \perp\!\!\!\perp Z_i \setminus \{z_i\}|z_i$. We consider several cases below:

Case 1. If $Z_1 \cap {}_a\pi_c \neq \emptyset$, then $a \perp\!\!\!\perp c|Z_1$, $\rho^2_{ac|Z_1} = 0$ and the result is trivial.

We initially assume that Z_1 separates Z_2 from a and c. This implies that for each $z_2 \in Z_2$ there is a $z^a_1 \in Z_1$ and $z^c_1 \in Z_1$ such that $z^a_1 \in {}_a\pi_{z_2}$ and $z^c_1 \in {}_c\pi_{z_2}$.

Case 2. If $Z_2 \cap {}_a\pi_c \neq \emptyset$, then $z_1^a \in {}_a\pi_{z_2} \subseteq {}_a\pi_c$. This implies that $a \perp\!\!\!\perp c|Z_1$ and $\rho^2_{ac|Z_1} = 0$.

Case 3. Now let $(Z_1 \cup Z_2) \cap {}_a\pi_c = \emptyset$. Suppose $Z_1 = \{z_{11}, z_{12}, \ldots, z_{1n_1}\}$ and $Z_2 = \{z_{21}, z_{22}, \ldots, z_{2n_2}\}$. Suppose $x_i = {}_a\pi_{z_{1i}} \cap {}_c\pi_{z_{1i}} \cap {}_a\pi_c$. Since G is a tree x_i is unique for z_i. Also suppose that $N_i = \{z_{2i} \in Z_2 \ : \ z_{1i} \in {}_a\pi_{z_{2i}} \cap {}_c\pi_{z_{2i}}\}$. Again from the structure of G it is clear that N_i are disjoint and $Z_2 = \cup_{i=1}^{n_1} N_i$. We don't exclude the possibility that N_i may be $\emptyset$ for some i. Using (2) we can write:

$$\frac{\rho^2_{ac|Z_1}}{\rho^2_{ac|Z_2}} = \prod_{i=1}^{n_1} \frac{\rho^2_{ac|z_{11}\ldots z_{1(i-1)}z_{1i}N_{i+1}\ldots N_{n_1}}}{\rho^2_{ac|z_{11}\ldots z_{1(i-1)}N_i N_{i+1}\ldots N_{n_1}}}. \tag{9}$$

It is sufficient to show that each factor in the product (9) is bounded by 1. Consider the ith factor,

$$f_i = \frac{\rho^2_{ac|z_{11}\ldots z_{1(i-1)}z_{1i}N_{i+1}\ldots N_{n_1}}}{\rho^2_{ac|z_{11}\ldots z_{1(i-1)}N_i N_{i+1}\ldots N_{n_1}}}.$$

Notice that the factor f_i depends only on the subgraph G_{V_i} of G defined by the vertex set:

$$V_i = \left\{ \bigcup_{j=1}^{i-1} \left({}_a\pi_{z_{1j}} \cup {}_c\pi_{z_{1j}} \right) \right\} \bigcup \left\{ \bigcup_{j=i}^{n_1} \bigcup_{z_{2k}^{(j)} \in N_j} \left({}_a\pi_{z_{2k}^{(j)}} \cup {}_c\pi_{z_{2k}^{(j)}} \right) \right\}.$$

It is clear that, G_{V_i} is a tree. Let us denote $B_i = \{z_{11}, \ldots, z_{1(i-1)}\} \cup (\cup_{j=i+1}^{n_1} N_j)$ and $B_i^c = V_i \setminus B_i$.

Now from the structure of G_{V_i} we note that (i) $x_i \in {}_a\pi_c$ so $a \perp\!\!\!\perp c|x_i B_i$, (ii) $x_i \in {}_a\pi_{z_{1i}}$ and $x_i \in {}_c\pi_{z_{1i}}$ implying $a \perp\!\!\!\perp z_{1i}|x_i, B_i$ and (iii) $z_{1i} \in \bigcup_{z_{2k}^{(i)} \in N_i} ({}_a\pi_{z_{2k}^{(i)}} \cap {}_c\pi_{z_{2k}^{(i)}})$ it follows that $ac \perp\!\!\!\perp N_i|z_{1i} B_i$.

From Lemma 1 it follows that the triples $\langle a, c \mid x_i \rangle$, $\langle ac, z_{1i} \mid x_i \rangle$ and $\langle ac, N_i \mid z_{1i} \rangle$ are in $\mathfrak{I}(G) \left[^{B_i} = \mathfrak{I}(G_{B_i^c})\right.$. It is obvious that,

$$\frac{\rho^2_{ac|B_i z_{1i}}(\Sigma)}{\rho^2_{ac|B_i N_i}(\Sigma)} = \frac{\rho^2_{ac|z_{1i}}\left(\Sigma_{B_i^c B_i^c|B_i}\right)}{\rho^2_{ac|N_i}\left(\Sigma_{B_i^c B_i^c|B_i}\right)}.$$

Now consider the following sub-cases:

(a) If $N_i = \emptyset$ or $N_i^{(1)} = z_{2i}$, from the Theorem 2 it follows that $\rho^2_{ac|z_{1i}}(\Sigma_{B_i^c B_i^c|B_i}) \leq \rho^2_{ac|N_i}(\Sigma_{B_i^c B_i^c|B_i})$.

(b) If $N_i = \{z_{21}, \dots, z_{2m_i}\}$, then using $ac \perp\!\!\!\perp N_i | z_{1i}$, we can write:

$$f_i = \frac{\rho^2_{ac|z_{1i}z_{22}\dots z_{2m_i}} \left(\Sigma_{B_i^c B_i^c|B_i}\right)}{\rho^2_{ac|z_{21}z_{22}\dots z_{2m_i}} \left(\Sigma_{B_i^c B_i^c|B_i}\right)}.$$

By following the same argument as above and conditioning on $\{z_{22}, \dots, z_{2m_i}\}$ it follows that $f_i \le 1$.

Now suppose that there is a $Z_2' \subseteq Z_2$ s.t. Z_2' is not separated from a and c by Z_1, but because of the choice of parameters both $\rho^2_{aZ_2'|Z_1} = \rho^2_{cZ_2'|Z_1} = 0$.

It can be shown that $\rho^2_{ac|(Z_2' \cup Z_1)} = \rho^2_{ac|Z_1}$. So if $Z_2' \cap {}_a\pi_c \neq \emptyset$ then $\rho^2_{ac|Z_2} = \rho^2_{ac|(Z_2' \cup Z_1)} = \rho^2_{ac|Z_1} = 0$. On the other hand if $Z_2' \cap {}_a\pi_c = \emptyset$ we can write:

$$\frac{\rho^2_{ac|Z_1}}{\rho^2_{ac|Z_2}} = \frac{\rho^2_{ac|(Z_2' \cup Z_1)}}{\rho^2_{ac|(Z_2' \cup \{Z_2 \setminus Z_2'\})}}. \tag{10}$$

The fact that the ratio in (10) is less than 1 follows from the first part mutatis mutandis. □

Proof of Corollary 2 The assumptions imply that Z_1 separates Z_2 from a and c. This is exactly Case 3. in the previous proof. □

Proof of Theorem 8 We parametrise the Cholesky decomposition $\Lambda = BB^{\mathrm{T}}$.

Suppose $z_1 \in Z_1$ and $z_2 \in Z_2$ such that $ac \not\perp\!\!\!\perp z_1 | Z_2$ and $ac \not\perp\!\!\!\perp z_2 | Z_1$. Let ${}_a\pi_c = \{a = v_1, v_2, \dots, v_d = c\}$, ${}_a\pi_c \cap {}_a\pi_{z_1} \cap {}_c\pi_{z_1} = v_i$, ${}_a\pi_c \cap {}_a\pi_{z_2} \cap {}_c\pi_{z_2} = v_j$, $i, j \in \{1, 2, \dots, d\}$. Further let ${}_{v_i}\pi_{z_1} = \{v_i, x_1, \dots, x_{d_1} = z_1\}$ and ${}_{v_i}\pi_{z_2} = \{v_j, y_1, \dots, y_{d_2} = z_2\}$. If $i = j$ it is possible that ${}_{v_i}\pi_{z_1}$ and ${}_{v_i}\pi_{z_2}$ intersect at more than one vertex. However, it does not change the proof, so w.l.g. we assume that $i \neq j$. Suppose

$$\begin{aligned} V_I &= {}_a\pi_c \cup {}_{v_i}\pi_{z_1} \cup {}_{v_j}\pi_{z_2} \\ E_I &= \{(v_2, v_1), \dots, (v_d, v_{d-1}), (x_1, v_i), \dots, (z_1, x_{d_1-1}), (y_1, v_j), \dots, (z_2, y_{d_2-1})\}. \end{aligned}$$

We list the variables in Σ as ${}_a\pi_c$, ${}_{v_1}\pi_{z_1}$, ${}_{v_2}\pi_{z_2}$, $V \setminus V_I$, where the vertices in $V \setminus V_I$ can be arranged in an arbitrary fashion. The matrix B inherits the same arrangement.

The matrix B is given by, $B_{kl} = 1$, {if $k = l$}, $B_{kl} = -1$, {if $(k, l) \in E_I$}, $B_{kl} = -b_1$, {if $(k, l) = (z_1, x_{d_1-1})$}, $B_{kl} = -b_2$, {if $(k, l) = (z_2, x_{d_2-1})$}, $B_{kl} = 0$, {otherwise}.

It can be shown that the resulting Λ is a n.n.d. matrix for all values of b_1 and b_2 and will represent all the conditional independence relations on the tree under consideration.

Now choose $b_1 = 0$. This implies $\rho^2_{ac|Z_1} = \rho^2_{ac} \ge \rho^2_{ac|z_2} = \rho^2_{ac|Z_2}$. The opposite happens if $b_2 = 0$. This completes the proof. □

Proof of Corollary 3 The result is trivial if ${}_a\pi_{c'} \cap Z' \neq \emptyset$. Furthermore, by assumption if Z intersects ${}_a\pi_c$, so does Z'. The non-trivial case can be shown by applying Theorem 6 and Corollary 2 respectively on the factors below:

$$\frac{\rho^2_{ac'|Z'}}{\rho^2_{ac|Z}} = \frac{\rho^2_{ac'|Z'}}{\rho^2_{ac|Z'}} \frac{\rho^2_{ac|Z'}}{\rho^2_{ac|Z}}.$$

□

To prove Theorem 9 we need the following definitions from the literature of DAGs.

Definition 3 A vertex v on a path ${}_x\pi_y$ in a polytree is a collider on the path if there are vertices v_1 and v_2 on ${}_x\pi_y$ such that the edges $v_1 \to v$ and $v_2 \to v$ exist. A vertex on a path ${}_x\pi_y$ in a polytree is a non-collider on the path if it is not a collider on ${}_x\pi_y$.

Definition 4 (*d-connection*) A path ${}_x\pi_y$ between x and y in a DAG is said to be d-connecting given a set Z (possibly empty) if 1. every non-collider on ${}_x\pi_y$ is not in Z and 2. every collider on ${}_x\pi_y$ is in $\mathrm{an}(Z)$. Here $\mathrm{an}(Z) = \cup_{z\in Z}\mathrm{an}(z)$.

If there is no path d-connecting x and y given Z, then x and y are said to be d-separated given Z.

Definition 5 For disjoint sets X, Y, Z, where Z may be empty, X and Y are d-separated given Z, if for every pair x, y, with $x \in X$ and $y \in Y$, x and y are d-separated given Z.

Definition 6 We say a density f factors according to a DAG, if for three disjoint sets X, Y and Z, $X \perp\!\!\!\perp Y|Z$ according to f whenever X is d-separated from Y given Z.

Proof of Theorem 9 First of all note that, since $\rho^2_{ac|bz} \neq \rho^2_{ac|b}$, $ac \not\perp\!\!\!\perp z \mid b$. Further, since $a \neq c \neq b$, $a \in \mathrm{an}(c)$ and $c \in \mathrm{an}(b)$, there are no colliders on ${}_a\pi_b$. We first show that $z \perp\!\!\!\perp ac$ iff $z \perp\!\!\!\perp c$. Clearly, $z \perp\!\!\!\perp ac$ implies $z \perp\!\!\!\perp c$. To show the converse first note that, since the graph is a polytree, if $z \perp\!\!\!\perp c$ there is at least one collider v on the unique path ${}_c\pi_z$ between c and z. Clearly, v cannot be on ${}_a\pi_c$, otherwise it will be a collider on ${}_a\pi_c$. However, by construction ${}_c\pi_z \setminus {}_a\pi_c = ({}_a\pi_z \cap {}_c\pi_z) \setminus {}_a\pi_c$. So if v is not on ${}_a\pi_c$, v would be a collider on ${}_a\pi_z$ as well. Thus, using the assumption that the graph is a polytree, $a \perp\!\!\!\perp z$ and our claim follows.

Similar argument shows if $z \perp\!\!\!\perp b$ iff $z \perp\!\!\!\perp acb$. So, $\rho^2_{ac|bz} \neq \rho^2_{ac|b}$ implies that $z \not\perp\!\!\!\perp b$. So only the following three cases, (i) $a \perp\!\!\!\perp z$ and $c \not\perp\!\!\!\perp z$, (ii) $ac \perp\!\!\!\perp z$ (i.e. $z \perp\!\!\!\perp c$) and (iii) $a \not\perp\!\!\!\perp z$ and $c \not\perp\!\!\!\perp z$ are possible (see Fig. 9). We first consider the if parts:

Case (i) We show that there is a vertex v_1 such that $az \perp\!\!\!\perp cb|v_1$. $a \perp\!\!\!\perp z$ implies there is at least one collider v_1 on ${}_a\pi_z$, $a \neq z \neq v_1$. Again by construction ${}_c\pi_{v_1} \setminus {}_a\pi_c = ({}_a\pi_{v_1} \cap {}_c\pi_{v_1}) \setminus {}_a\pi_c$. Thus, if $v_1 \notin {}_a\pi_c$, v_1 is a collider on ${}_c\pi_z$ as well, which would imply $c \perp\!\!\!\perp z$. Thus $v_1 \in {}_a\pi_c$. Clearly, v_1 cannot be a collider on ${}_a\pi_c$. Thus v_1 is the only collider on ${}_a\pi_z$ and it is not a collider on ${}_a\pi_c$ and ${}_c\pi_z$. Thus, from the definition of d-separation it follows that $az \perp\!\!\!\perp cb|v_1$. From Theorem 4(ii) it follows that $\rho^2_{ac|bz} > \rho^2_{ac|b}$.

Case (ii) We show that $a \perp\!\!\!\perp z|cb$ and apply Theorem 3 with $x = c$. Since by assumption $c \perp\!\!\!\perp z$ and $b \not\perp\!\!\!\perp z$, as in Case (i) above there is a vertex $v_2 \in {}_c\pi_b$ such that

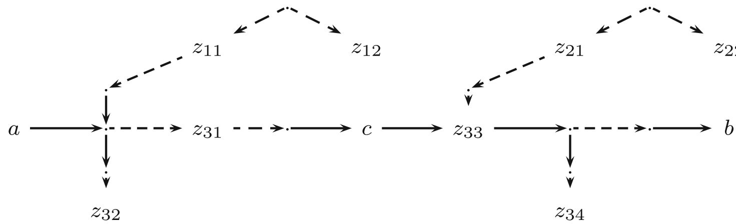


Fig. 9 Example of a polytree discussed in Theorem 9. Vertices z_{11} and z_{12} are relevant to Case (i), z_{21} and z_{22} are relevant to Case (ii) below. The vertices z_{3k}, for $k = 1, \ldots, 4$ corresponds to Case (iii) in the proof below

v_2 is a collider on ${}_c\pi_z$ but not a collider on ${}_b\pi_z$. Note that, $v_2 \neq c$ or $v_2 \neq z$. Thus c is a non-collider on both ${}_a\pi_z$ and ${}_a\pi_b$ and c d-separates a from $\{b, z\}$. This implies $a \perp\!\!\!\perp bz|c$, which in turn gives $a \perp\!\!\!\perp z|cb$. Now from Theorem 3 we get $\rho^2_{ac|bz} < \rho^2_{ac|b}$.

Case (iii) Since $a \not\perp\!\!\!\perp z$, it follows that $c \not\perp\!\!\!\perp z$ and $b \not\perp\!\!\!\perp z$. This implies there is no collider on ${}_a\pi_z$, ${}_c\pi_z$ and ${}_b\pi_z$. Let $v_3 = {}_a\pi_z \cap {}_c\pi_z \cap {}_b\pi_z \cap {}_a\pi_b$. Clearly, v_3 is a non-collider on all these paths. So, it follows that $acb \perp\!\!\!\perp z|v_3$ (Lauritzen 1996, page 29). This implies $ac \perp\!\!\!\perp z|bv_3$. Further, if $v_3 \in {}_a\pi_c$, $a \perp\!\!\!\perp cb|v_3$ and $a \perp\!\!\!\perp c|bv_3$. It is possible that $z = v_3$. Now if $v_3 \in {}_a\pi_c$, Theorem 3 with $x = v_3$ imply $\rho^2_{ac|bz} < \rho^2_{ac|b}$. Note that in this case if $v_3 = z$, $\rho^2_{ac|bz} = 0$. If $v_3 \notin {}_a\pi_c$, we consider two cases. Case (a) $z = v_3 \in {}_c\pi_b$. Clearly $ac \perp\!\!\!\perp b|z$. Now using Theorem 3 we get $\rho^2_{ac|bz} = \rho^2_{ac|z} < \rho^2_{ac|b}$. Case (b) When $v_3 \notin {}_c\pi_b$ use Theorem 3 on conditional covariance given b with $x = c$ to get $\rho^2_{ac|bz} < \rho^2_{ac|b}$.

The 'only if' parts follow from the 'if' part and the fact that the above three are only possible cases under our assumptions. □

Acknowledgments The author would like to thank Michael Perlman, Thomas Richardson, Mathias Drton, Antar Bandyopadhyay, the referees and the associate editor for their useful comments and suggestions during the preparation of this article.

References

Bekker, P. A., de Leeuw, J. (1987). The rank of reduced dispersion matrices. *Psychometrika*, *52*, 125–135.

Chaudhuri, S. (2005). *Using the structure of d-connecting paths as a qualitative measure of the strength of dependence*. PhD thesis, Seattle: Department of Statistics, University of Washington.

Chaudhuri, S. (2013). *Qualitative inequalities for squared partial correlations of a gaussian random vector*. Technical Report 1/2013, Department of Statistics and Applied Probability, National University of Singapore.

Chaudhuri, S., Richardson, T. S. (2003). Using the structure of d-connecting paths as a qualitative measure of the strength of dependence. In: *Proceedings of the Nineteenth Conference Conference on Uncertainty in Artificial Intelligence* (pp. 116–123). Morgan Kaufmann, San Francisco, CA

Chaudhuri, S., Tan, G. L. (2010). On qualitative comparison of partial regression coefficients for gaussian graphical Markov models. In M. A. G. Viana, H.P. Wynn (Eds.), *Algebraic methods in statistics and probability II. Contemporary mathematics*, Vol. 516 (pp. 125–133). Providence, Rhode Island: American Mathematical Society.

Cheng, J., Greiner, R., Kelly, J., Bell, D., Liu, W. (2002). Learning bayesian networks from data: an information-theory based approach. *Artificial Intelligence*, *137*, 43–90.

Chickering, D., Meek, C. (2006). On the compatibility of faithfulness and monotone dag faithfulness. *Artificial Intelligence*, *170*, 653–666.

Cover, T., Thomas, J. (2006). *Elements of Information Theory*. Hoboken: Wiley

Drton, M., Strumfels, B., Sullivant, S. (2007). Algebraic factor analysis: tetrads, pentads and beyond. *Probability and related fields*, *138*, 463–493.

Greenland, S. (2003). Quantifying biases in causal models: classical confounding versus collider-stratification bias. *Epidemiology*, *14*, 300–306.

Greenland, S., Pearl, J. (2011). Adjustments and their consequences-collapsibility analysis using graphical models. *International Statistical Review*, *79*(3), 401–426.

Jarvie, H. P., Neal, C., Withers, P. J., Wescott, C., Acornley, R. M. (2005) Nutrient hydrochemistry for a groundwater-dominated catchment: the Hampshire Avon, UK. *Science of The Total Environment*, *344* (1–3), 143–158.

Kuroki, M., Cai, Z. (2006). On recovering a population covariance matrix in the presence of selection bias. *Biometrika*, *93*(3), 601–611.

Lauritzen, S. (1996). *Graphical models*. Oxford: Oxford University Press, Inc.

Lněnička, R., Matúš, F. (2007). On Gaussian conditional independence structures. *Kybernetika*, *43*(3), 327–342.

Matúš, F. (2005). Conditional independence in gaussian vectors and rings of polynomials. In G. Kern-Isberner, W. Rödder, F. Kulmann (Eds.), *Conditionals, information, and inference (WCII 2002 Hagen)* (pp. 152–161). Berlin: Springer.

Matúš, F. (2006). Piecewise linear conditional information inequality. *IEEE Transaction on Information Theory*, *52*(1), 236–238.

Matúš, F. (2007). Infinitely many information inequalities. In IEEE international symposium on information theory, 2007. ISIT 2007 (pp. 41–44).

Richardson, T., Spirtes, P. (2002). Ancestral graph markov models. *The Annals of Statistics*, *30*(4), 962–1030.

Roberts, G. O., Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society Series B Methodological*, *59*(2), 291–317.

Rodríguez-Iturbe, I., Rinaldo, A. (2001). *Fractal River Basins chance and self-organisation*. Cambridge: Cambridge University Press.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, *7*, 2003–2030.

Spirtes, P., Glymour, C., Scheines, R. (2000). *Causation, prediction, and search*. Cambridge: MIT Press.

VanderWeele, T. J., Robins, J. M. (2007). Directed acyclic graphs, sufficient causes, and the properties of confounding on a common effect. *American Journal of Epidemiology*, *166*(9), 1096–1104.

VanderWeele, T. J., Robins, J. M. (2010). Signed directed acyclic graphs for causal inference. *Journal of the Royal Statistical Society: Series B Methodological*, *72*(1), 111–127.

Verma, T., Pearl, J. (1990). Equivalence and synthesis of causal models. In P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), *Proceedings of the Sixth Conference Conference on Uncertainty in Artificial Intelligence* (pp. 220–227). Corvallis: AUAI Press.

Wermuth, N., Cox, D. R. (2008). Distortion of effects caused by indirect confounding. *Biometrika*, *95*(1), 17–33.

Whittaker, J. (2008). *Graphical models in applied multivariate statistics*. Chichester: Wiley.

Xu, L., Pearl, J. (1989). Structuring causal tree models with continuous variables. In M. Henrion, R. Shachter, L. Kanal, J. Lemmer (Eds.), *Proceedings of the fifth conference on uncertainty in artificial intelligence* (pp. 170–178). Corvallis: AUAI Press.

Zhang, Z., Yeung, R. W. (1997). A non-shannon-type conditional inequality of information quantities. *IEEE Transaction on Information Theory*, *43*(6), 1982–1986.