
Using the structure of d-connecting paths as a qualitative measure of the strength of dependence.

Sanjay Chaudhuri
Department of Statistics
University of Washington
Seattle, WA 98195

Thomas Richardson
Department of Statistics
University of Washington
Seattle, WA 98195

Abstract

Pearl’s concept of a d-connecting path is one of the foundations of the modern theory of graphical models: the absence of a d-connecting path in a DAG indicates that conditional independence will hold in any distribution factorizing according to that graph. In this paper we show that in singly-connected Gaussian DAGs it is possible to use the form of a d-connecting path to obtain qualitative information about the strength of conditional dependence. More precisely, the squared partial correlations between two given variables, conditioned on different subsets may be partially ordered by examining the relationship between the d-connecting path and the set of variables conditioned upon.

1 Introduction

Central to the modern theory of graphical models is the concept of d-separation which provides a simple algorithm for determining which conditional independence relations will hold in a distribution factorizing according to a DAG. Completeness results [4], [7] show further that whenever a d-connecting path exists between x and y given Z in a DAG G then in ‘almost all’ distributions that factor according to G , x and y will be dependent given Z .

However, recent work has shown that not all d-connecting paths are created equal: Greenland [5] shows that in certain specific contexts relevant to causal analysis in Epidemiology, shorter d-connecting paths lead to qualitatively stronger dependence than longer paths.

In this paper we analyse the relationship between the strength of dependence resulting from a single

d-connecting path and the set of variables which have been conditioned on in the situation where the joint distribution is Gaussian and the DAG is singly-connected.

Since colliders are made ‘active’ by conditioning on their descendants it seems as if the strength of dependence ought to be inversely related to the length of the path from a collider to a vertex in the conditioning set. Likewise one would expect dependence to decrease if we condition on vertices ‘near’ to a non-collider on a path. Such intuitions are given additional impetus by the intuitive description of d-separation in terms of ‘causal pipes’ ([9], p.72). However, our analysis shows that though there are situations in which these intuitions are correct (Lemmas 3.1 and 3.2) there are also contexts where these intuitions are either incorrect or provide no guidance (Lemmas 3.3 and 3.4).

An understanding of the relationship between the graphical structure and strength of dependence will facilitate sensitivity analysis in causal inference and may suggest new strategies for constraint based search [3],[2].

The paper is organised as follows: Section 2 contains the basic definitions; in Section 3 we consider four canonical examples; in Section 4 we define a partial ordering on conditioning sets associated with a d-connecting path; Section 5 describes properties of singly connected DAGs; Section 6 contains the main result and Section 8 a brief discussion. Proofs of the results in Section 3 are in the Appendix.

2 Some Initial Definitions

Let $G = (V, E)$ be a DAG. A non-endpoint vertex ζ on a path is a *collider on the path* if the edges preceding and succeeding ζ on the path have an arrowhead at ζ , i.e. $\rightarrow \zeta \leftarrow$. A non-endpoint vertex ζ on a path which is not a collider is a *non-collider on the path*. A path between vertices α and β in a DAG G is said to be

d -connecting given a set Z (possibly empty) if

- (i) every non-collider on the path is not in Z , and
- (ii) every collider on the path is in $\text{an}(Z)$.

where $\text{an}(Z) = \{x \mid x \rightarrow \dots \rightarrow z \text{ or } x = z \text{ in } G\}$. If there is no path d -connecting α and β given Z , then α and β are said to be d -separated given Z . For disjoint sets X, Y, Z , where Z may be empty, X and Y are d -separated given Z , if for every pair α, β , with $\alpha \in X$ and $\beta \in Y$, α and β are d -separated given Z .

Let $V = \{v_1, \dots, v_n\}$. When the joint distribution is normal the model may be expressed by the following equations

$$v_i = \mathbf{b}_{v_i} \cdot \mathbf{pa}_{v_i} + \epsilon_{v_i} \quad (1)$$

where \mathbf{pa}_{v_i} is the vector of parents of v_i and \mathbf{b}_{v_i} is the vector of coefficients $b_{v_i v_j}$, $v_j \in \text{pa}(v_i)$. ϵ_{v_i} follows a $N(0, \tau_{v_i}^2)$ distribution. The joint variance-covariance matrix for V is given by :

$$\Sigma_V = B^{-1} \Delta (B^T)^{-1} \quad (2)$$

where

$$\Delta = \text{diag}(\tau_{v_1}^2, \tau_{v_2}^2, \dots, \tau_{v_n}^2) \quad (3)$$

$$B_{ij} = \begin{cases} 1 & \text{if } i = j \\ -b_{v_i v_j} & \text{if } v_j \in \text{pa}(v_i) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

We assume that $|b_{v_i v_j}| > 0$ if $v_j \in \text{pa}(v_i)$ and $\tau_{v_i}^2 > 0$ so Σ_V is positive definite .

We use $\rho_{AC|B}^2$ as our measure of association of A and C conditional on B . Note that it is a monotonic transformation of the information proper of the conditional independence of A and C given B . The *information proper* for normal models takes the form

$$\text{Inf}(A \perp\!\!\!\perp C|B) = -\frac{1}{2} \log \left(1 - \rho_{AC|B}^2 \right) \quad (5)$$

we cite [10] for details.

For vertices X, Y and a collection of vertices Z by $\sigma_{XY|Z}$ and $\sigma_{XX|Z}$ we denote respectively the conditional covariance between X and Y given Z and the conditional variance of X given Z . Note that Z may be \emptyset . If X and Y are each a collection of vertices, then Σ_{XY} will denote the $X \times Y$ sub-matrix of Σ_V in (2). Throughout this article we shall repeatedly use the expression for conditional covariance in the Gaussian case:

$$\sigma_{AB|C} = \frac{\sigma_{AB}\sigma_{CC} - \sigma_{AC}\sigma_{BC}}{\sigma_{CC}} \quad (6)$$

This easily gives the expression for the conditional correlation and regression coefficients (see [6]). More generally if X_1, X_2, \dots, X_p jointly follows a $N(\mu, \Sigma)$ distribution then the iterative expression

$$\sigma_{12|3\dots p} = \sigma_{12|3\dots(p-1)} - \frac{\sigma_{1p|3\dots(p-1)} \cdot \sigma_{p2|3\dots(p-1)}}{\sigma_{pp|3\dots(p-1)}} \quad (7)$$

holds. Similar iterative expressions exist for the conditional correlation and regression coefficients. (see [6](page 346) or [1]).

3 Canonical examples

We start by looking at some specific key cases. The motivation for this being the fact that later it will be shown that a more general case can be reduced to these. Proofs are in the appendix.

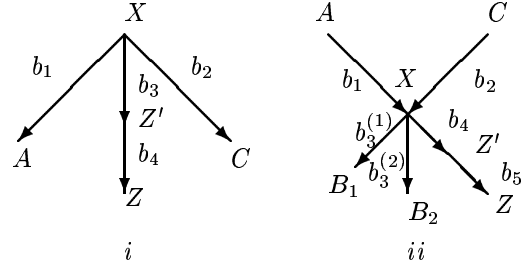


Figure 1: Cases where intuitions based on d -connection are correct. See Lemmas 3.1 and 3.2

Lemma 3.1. *In the graph in figure 1.i*

$$0 = \rho_{AC|X}^2 \leq \rho_{AC|Z'}^2 \leq \rho_{AC|Z}^2$$

holds, or in other words the squared correlation decreases with proximity to the path.

Note that this fits with the intuition that the closer we are to X , the more nearly we block the $A - C$ path. The result directly generalises to a graph in which $X \rightarrow \dots \rightarrow Z' \rightarrow \dots \rightarrow Z$.

Lemma 3.2. *In the graph in figure 1.ii*

$$\rho_{AC|BZ'}^2 \geq \rho_{AC|BZ}^2$$

holds, where B is a set of descendants of X or in other words the squared correlation increases with proximity to X . (In Fig 1.ii, $B = \{B_1, B_2\}$.)

Again this fits the intuition that the further we move from the path the weaker is our information about X and hence the closer we are to d -separating.

Lemma 3.3. *In the graph in figure 2.i*

$$\rho_{AC|BZ'}^2 \leq \rho_{AC|BZ}^2$$

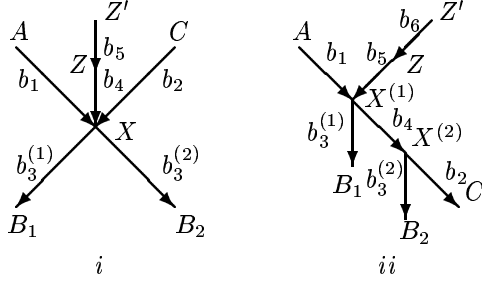


Figure 2: The anomalous cases.

holds, where $B = \{B_1, B_2\}$ or in other words the squared correlation increases with proximity to vertex X .

Lemma 3.4. In the graph in figure 2.ii

$$0 = \rho_{AC|BX^{(1)}}^2 \leq \rho_{AC|BZ'}^2 \leq \rho_{AC|BZ}^2$$

holds, where B is a set of descendants of $X^{(1)}$, or in other words the squared correlation increases with proximity to vertex $X^{(1)}$, though conditioning on $X^{(1)}$ d -separates A and C .

The following is an extension of the graph in figure 2.ii.

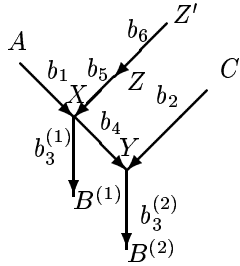


Figure 3: An extension of the case in figure 2.ii.

Lemma 3.5. In the graph in figure 3

$$0 = \rho_{AC|BX^{(1)}}^2 \leq \rho_{AC|BZ'}^2 \leq \rho_{AC|BZ}^2$$

holds, where B is a set of descendants of $X \cup Y$, or in other words the squared correlation increases with proximity to vertex X , though conditioning on X d -separates A and C .

In figure 4 we plot the squared correlation for the above cases as we increase the length of the path between X and Z . Figure 1.i is for the graph in figure 1.i. Figure 1.ii is for that in figure 1.ii, with $B = \emptyset$. Figure 2.i is for figure 2.i with $|B| = 1$ and the last figure in the plot is for figure 2.ii with $B = \emptyset$. All the parameter values are fixed at 1. In the plot the curved lines show

the values of the squared correlations. The broken lines denote ρ_{AC}^2 in case of figure 1.i, $\rho_{AC|X}^2$ in figure 1.ii, $\rho_{AC|B}^2$ in figure 2.i and ρ_{AC}^2 in figure 2.ii. In the Figure 2.i we also note that the value of $\rho_{AC|X}^2$ is larger than the $\rho_{AC|B}^2$ as expected from Lemma 3.2, also it seems that that the solid curve in this plot converges to $\rho_{AC|X}^2$.

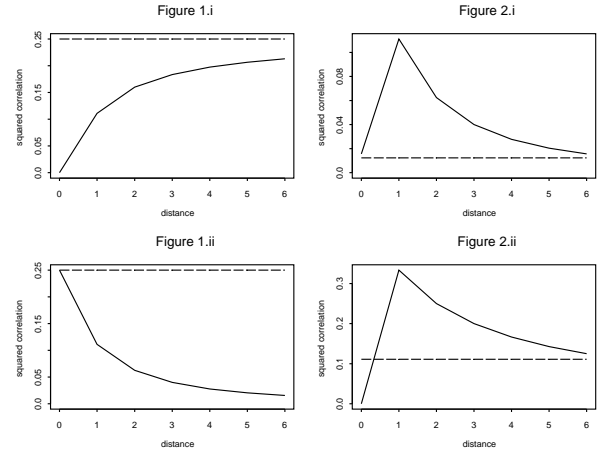


Figure 4: Squared correlation as a function of distance from the path to the vertex conditioned on (see text for explanation)

4 A partial ordering on conditioning sets associated with a path

Building on the previous examples we now construct a partial ordering on conditioning sets associated with a specific path. Let $G = (V, E)$ be a singly connected DAG i.e. the adjacencies in G form a tree. Let A and C be two nodes $A \neq C$. By a *path* ${}_A\pi_C$ we mean a sequence of non repeating adjacent nodes $\{x_1, x_2, \dots, x_n\} \subseteq V$ s.t. $x_1 = A$ and $x_n = C$. Since G is singly connected for any two vertices $A, C, A \neq C$ there is a unique ${}_A\pi_C$ connecting A and C .

Let

$$Z = \{z : \exists x \in {}_A\pi_C \text{ so that } x \text{ is } d\text{-connected to } z \text{ given the } \emptyset\} \quad (8)$$

Define 2^Z = the power set of Z . Below we shall use graphical structure to define a partial order on 2^Z . Let \mathbb{Z} denote an element in 2^Z .

Since G is tree structured then for any $z \in \mathbb{Z}$ there is a unique $x_z^* \in {}_A\pi_C$ so that ${}_A\pi_z \cap {}_C\pi_z \cap {}_A\pi_C = x_z^*$. For each $\mathbb{Z} \in 2^Z$ define

$$\mathbb{Z}_{nc} = \{z \in \mathbb{Z} : \text{none of } {}_A\pi_C, {}_C\pi_z, {}_A\pi_z \text{ has a collider at } x_z^*\} \quad (9)$$

$$\mathbb{Z}_c = \{z \in \mathbb{Z} : \text{at least one of } {}_A\pi_C, {}_C\pi_z, \quad (10)$$

$${}_A\pi_z \text{ has a collider at } x_z^*\}$$

We further subdivide \mathbb{Z}_c in two parts :

$$\mathbb{Z}_{c-c} = \{z \in \mathbb{Z} : \text{at least one of } {}_C\pi_z, {}_A\pi_z \quad (11)$$

$$\text{has a collider at } x_z^*\},$$

$$\mathbb{Z}_{c-nc} = \{z \in \mathbb{Z} : \text{none of } {}_C\pi_z, {}_A\pi_z \text{ has a} \quad (12)$$

$$\text{collider at } x_z^*, \text{ but } {}_A\pi_C$$

$$\text{has a collider at } x_z^*\}.$$

Note that in definitions (11) and (12) if for some $z \in \mathbb{Z}_c$, $z = x_z^*$, then since ${}_A\pi_z$ and ${}_C\pi_z$ do not have colliders at z , (endpoints are neither colliders nor non-colliders) there is a collider at z in ${}_A\pi_C$. So $z \in \mathbb{Z}_{c-nc}$.

Let $z \in \mathbb{Z}_{nc}$, then define z to be a *nearest vertex* to ${}_A\pi_C$ in \mathbb{Z}_{nc} if there is no $z' \neq z \in \mathbb{Z}_{nc}$ s.t. $z' \in {}_A\pi_z \cap {}_C\pi_z$. Let

$$N(\mathbb{Z}_{nc}) = \text{set of all nearest nodes to } {}_A\pi_C \text{ in } \mathbb{Z}_{nc}. \quad (13)$$

Similarly we define :

$$N(\mathbb{Z}_{c-c}) = \text{set of all nearest nodes to } {}_A\pi_C \text{ in } \mathbb{Z}_{c-c},$$

$$N(\mathbb{Z}_{c-nc}) = \text{set of all nearest nodes to } {}_A\pi_C \text{ in } \mathbb{Z}_{c-nc}.$$

Let $S_1, S_2 \in 2^Z$. Then (S_2, S_1) is a *total further - nearer pair* w.r.t. ${}_A\pi_C$ if

- For each $z_2 \in S_2$, $\exists z_1 \in S_1$ s.t. $z_1 \in {}_A\pi_{z_2} \cap {}_C\pi_{z_2}$.
- For each $z_1 \in S_1$, $\exists z_2 \in S_2$ s.t. $z_1 \in {}_A\pi_{z_2} \cap {}_C\pi_{z_2}$.

Let $\mathbb{Z}^{(1)}, \mathbb{Z}^{(2)} \in 2^Z$. We define $\mathbb{Z}^{(1)} \prec \mathbb{Z}^{(2)} \in 2^Z$ if all the following are satisfied :

- $(N(\mathbb{Z}_{nc}^{(2)}), N(\mathbb{Z}_{nc}^{(1)}))$ is a total further - nearer pair w.r.t. ${}_A\pi_C$.
- $(N(\mathbb{Z}_{c-c}^{(1)}), N(\mathbb{Z}_{c-c}^{(2)}))$ is a total further - nearer pair w.r.t. ${}_A\pi_C$.
- $(N(\mathbb{Z}_{c-nc}^{(1)}), N(\mathbb{Z}_{c-nc}^{(2)}))$ is a total further - nearer pair w.r.t. ${}_A\pi_C$.

5 Some facts about singly connected DAGs

As before let $G = (V, E)$ be a singly connected DAG and ${}_A\pi_C$ be a path between two nodes $A \neq C$.

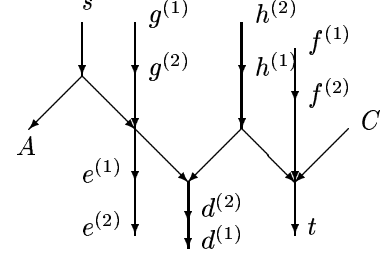


Figure 5: Example of two ordered conditioning sets: $\mathbb{Z}^{(1)} = \{s, g^{(1)}, e^{(1)}, d^{(1)}, h^{(1)}, f^{(1)}, t\} \prec \mathbb{Z}^{(2)} = \{s, g^{(2)}, e^{(2)}, d^{(2)}, h^{(2)}, f^{(2)}, t\}$. Here $\mathbb{Z}_{nc}^{(1)} = \{e^{(1)}, h^{(1)}, s\}$, $\mathbb{Z}_{nc}^{(2)} = \{e^{(2)}, h^{(2)}, s\}$, $\mathbb{Z}_{c-c}^{(1)} = \{g^{(1)}, f^{(1)}\}$, $\mathbb{Z}_{c-c}^{(2)} = \{g^{(2)}, f^{(2)}\}$, $\mathbb{Z}_{c-nc}^{(1)} = \{d^{(1)}, t\}$, $\mathbb{Z}_{c-nc}^{(2)} = \{d^{(2)}, t\}$.

Lemma 5.1. *For any $y, z \in V$, $y \neq z$, $|pa(y) \cap pa(z)| \leq 1$ and $|ch(y) \cap ch(z)| \leq 1$. Moreover at most one of $pa(y) \cap pa(z)$ and $ch(y) \cap ch(z)$ is nonempty. Also $\forall x \in V$, $x \notin {}_y\pi_z$, then there exists at most one $w \in {}_y\pi_z$ so that $x \in pa(w)$ or $x \in ch(w)$.*

Proof. If there are two common parents say x, x^* then the skeleton of the subgraph $G_{\{x, x^*, y, z\}}$ has a 4 cycle, which violates the assumption. Similar proof follows for common children.

The second part follows from the fact that if both of $pa(y) \cap pa(z)$ and $ch(y) \cap ch(z)$ are non empty then the skeleton of G has a 4 cycle.

Similar arguments prove the other statements. \square

Lemma 5.2. *Let $G = (V, E)$ be a singly connected DAG. Let G^* be the induced subgraph consisting of ${}_y\pi_z$, ($y \neq z$). Then for disjoint subsets S_1, S_2, S_3 occurring on ${}_y\pi_z$, S_1 is d -separated from S_2 given S_3 in G^* iff S_1 is d -separated from S_2 given S_3 in G .*

Proof. The proof is intuitively clear. For a detailed proof see Theorem 4.18 in [8]. \square

Lemma 5.3. *Let $G = (V, E)$ be a singly connected DAG. Then if $Z \subseteq V$, there exists a DAG $G^* = (V \setminus Z, E^*)$ such that for disjoint $S_1, S_2, S_3 \subseteq V \setminus Z$ (S_3 may be empty) S_1 and S_2 are d -separated given $S_3 \cup Z$ in G iff S_1 and S_2 are d -separated given S_3 in G^* .*

Proof. The proof follows from Theorem 4.18 and Theorem 4.2 in [8]. \square

In other words given a singly connected DAG G and a set Z we may always find a new DAG G^* which represents the d -separation relations holding in G after conditioning on Z .

6 The main theorems

Let $G = (V, E)$ be a singly connected DAG, $A, C \in V$, $A \neq C$. Let $\mathbb{Z}^{(1)}, \mathbb{Z}^{(2)} \in 2^Z$, so that $\mathbb{Z}^{(1)} \prec \mathbb{Z}^{(2)}$. The main result of this section is that $\rho_{AC|\mathbb{Z}^{(1)}}^2 \leq \rho_{AC|\mathbb{Z}^{(2)}}^2$.

Theorem 1. *If $\rho_{AC|\mathbb{Z}^{(2)}}^2 = 0$ then $\rho_{AC|\mathbb{Z}^{(1)}}^2 = 0$.*

Proof. Let $\rho_{AC|\mathbb{Z}^{(2)}}^2 = 0$. Since by assumption $|b_{v_i pa(v_i)}| > 0$ and $\tau_{v_i}^2 > 0$, $\forall v_i \in V$ it follows that A and C are d-separated given $\mathbb{Z}^{(2)}$. Then either there is a non-collider x on the path which is in $\mathbb{Z}^{(2)}$ or there is a collider x which has no descendants in $\mathbb{Z}^{(2)}$.

Case 1. If x is a non-collider on $A\pi_C$ then since x is an endpoint of $A\pi_x$ and $C\pi_x$, $x \notin \mathbb{Z}_c^{(2)}$, hence $x \in \mathbb{Z}_{nc}^{(2)}$. Since $x^* = x$, $x \in N(\mathbb{Z}_{nc}^{(2)})$ and since $(N(\mathbb{Z}_{nc}^{(2)}), N(\mathbb{Z}_{nc}^{(1)}))$ is a total further-nearer pair, so by (a) of the definition there is some $y \in N(\mathbb{Z}_{nc}^{(1)})$ s.t. $y \in A\pi_x \cap C\pi_x$. Hence $x \in N(\mathbb{Z}_{nc}^{(1)}) \subseteq \mathbb{Z}_{nc}^{(1)}$, since the paths only intersect at x . Thus A and C are d-separated given $\mathbb{Z}^{(1)}$.

Case (2): If there is a collider x s.t. $\text{de}(x) \cap \mathbb{Z}^{(2)} = \emptyset$, then since the collider is in $A\pi_C$ then $\text{de}(x) \cap \mathbb{Z}_{c-c}^{(2)} = \emptyset$ by construction and $\text{de}(x) \cap \mathbb{Z}_{c-nc}^{(2)} = \emptyset$ by assumption. Thus $\text{de}(x) \cap N(\mathbb{Z}_{c-nc}) = \emptyset$.

Now as $(N(\mathbb{Z}_{c-nc}^{(1)}), N(\mathbb{Z}_{c-nc}^{(2)}))$ form a total further-nearer pair so by (b) of the definition there is no $\text{de}(x)$ in $\mathbb{Z}_{c-nc}^{(1)}$. Thus by single connectedness $\text{de}(x) \cap \mathbb{Z}^{(1)} = \emptyset$. Note that $\text{de}(x) \cap \mathbb{Z}_{c-c}^{(1)} = \emptyset$. So A and C are d-separated given $\mathbb{Z}^{(1)}$.

We have shown that in this case $\rho_{AC|\mathbb{Z}^{(1)}}^2 = 0$. \square

Assume $\rho_{AC|\mathbb{Z}^{(2)}}^2 > 0$. If $\mathbb{Z}^{(i)}$ contains a non-collider in $A\pi_C$ trivially $\rho_{AC|\mathbb{Z}^{(i)}}^2 = 0$, $i = 1, 2$. Thus we can assume that none of $\mathbb{Z}^{(1)}$ and $\mathbb{Z}^{(2)}$ contain any non-collider in $A\pi_C$.

Let $N(\mathbb{Z}) = N(\mathbb{Z}_{nc}) \cup N(\mathbb{Z}_{c-c}) \cup N(\mathbb{Z}_{c-nc})$. Since the graph is singly connected,

$$\{A, C\} \perp\!\!\!\perp \mathbb{Z} \setminus N(\mathbb{Z}) | N(\mathbb{Z}) \quad (14)$$

So it suffices to show that

$$\rho_{AC|N(\mathbb{Z}^{(1)})}^2 \leq \rho_{AC|N(\mathbb{Z}^{(2)})}^2 \quad (15)$$

Let $N(\mathbb{Z}_{nc}^{(i)}) = \{a_1^{(i)}, a_2^{(i)}, \dots, a_{k_i}^{(i)}\}$, $N(\mathbb{Z}_{c-c}^{(i)}) = \{b_1^{(i)}, b_2^{(i)}, \dots, b_{l_i}^{(i)}\}$, $N(\mathbb{Z}_{c-nc}^{(i)}) = \{c_1^{(i)}, c_2^{(i)}, \dots, c_{m_i}^{(i)}\}$, $i = 1, 2$.

First consider the case where $k_1 = k_2 = k$, $l_1 = l_2 = l$ and $m_1 = m_2 = m$.

$$\begin{aligned} \frac{\rho_{AC|N_{\mathbb{Z}^{(1)}}}^2}{\rho_{AC|N_{\mathbb{Z}^{(2)}}}^2} &= \frac{\rho_{AC|\{a_1^{(1)}, \dots, a_k^{(1)}, b_1^{(1)}, \dots, b_l^{(1)}, c_1^{(1)}, \dots, c_m^{(1)}\}}^2}{\rho_{AC|\{a_1^{(2)}, \dots, a_k^{(2)}, b_1^{(2)}, \dots, b_l^{(2)}, c_1^{(2)}, \dots, c_m^{(2)}\}}^2} \\ &= \prod_{i=1}^k \frac{\rho_{AC|\{a_1^{(2)}, \dots, a_{i-1}^{(2)}, a_i^{(1)}, a_{i+1}^{(2)}, \dots, a_k^{(1)}, b_1^{(1)}, \dots, b_l^{(1)}, c_1^{(1)}, \dots, c_m^{(1)}\}}^2}{\rho_{AC|\{a_1^{(2)}, \dots, a_{i-1}^{(2)}, a_i^{(2)}, a_{i+1}^{(2)}, \dots, a_k^{(2)}, b_1^{(1)}, \dots, b_l^{(1)}, c_1^{(1)}, \dots, c_m^{(1)}\}}^2} \\ &\quad \times \prod_{i=1}^l \frac{\rho_{AC|\{a_1^{(2)}, \dots, a_k^{(2)}, b_1^{(2)}, \dots, b_{i-1}^{(2)}, b_i^{(1)}, b_{i+1}^{(2)}, \dots, b_l^{(1)}, c_1^{(1)}, \dots, c_m^{(1)}\}}^2}{\rho_{AC|\{a_1^{(2)}, \dots, a_k^{(2)}, b_1^{(2)}, \dots, b_{i-1}^{(2)}, b_i^{(2)}, b_{i+1}^{(2)}, \dots, b_l^{(1)}, c_1^{(1)}, \dots, c_m^{(1)}\}}^2} \\ &\quad \times \prod_{i=1}^m \frac{\rho_{AC|\{a_1^{(2)}, \dots, a_k^{(2)}, b_1^{(2)}, \dots, b_l^{(2)}, c_1^{(2)}, \dots, c_{i-1}^{(2)}, c_i^{(1)}, c_{i+1}^{(2)}, \dots, c_m^{(1)}\}}^2}{\rho_{AC|\{a_1^{(2)}, \dots, a_k^{(2)}, b_1^{(2)}, \dots, b_l^{(2)}, c_1^{(2)}, \dots, c_{i-1}^{(2)}, c_i^{(2)}, c_{i+1}^{(2)}, \dots, c_m^{(1)}\}}^2} \end{aligned} \quad (16)$$

The following theorems are proved by showing that each term in the above product is bounded by 1.

Theorem 2. *Let $G = (V, E)$ be a singly connected DAG. $A, C \in V$ and $A \neq C$. Suppose $A\pi_C$ does not have a collider. Let $\mathbb{Z}^{(1)} \prec \mathbb{Z}^{(2)}$ and $k_1 = k_2$, $l_1 = l_2$. Then*

$$\rho_{AC|\mathbb{Z}^{(1)}}^2 \leq \rho_{AC|\mathbb{Z}^{(2)}}^2$$

Proof. (Sketch) Note that since $A\pi_C$ does not have a collider then both $\mathbb{Z}_{c-nc}^{(1)}$ and $\mathbb{Z}_{c-nc}^{(2)}$ are empty. So the product (16) will only have the first two products, more over it will only involve the a and b nodes.

By choosing the nodes to be conditioned on and marginalised out in an appropriate order one can show that it is possible to reduce the graph corresponding to each term in the first product in (16) to a graph Markov equivalent to figure 1.i and that in the second product to that in figure 2.ii.

So by Lemma 3.1 and 3.4 the proof follows. \square

Theorem 3. *Let $G = (V, E)$ be a singly connected DAG. $A, C \in V$ and $A \neq C$. Suppose $A\pi_C$ has exactly one collider. Let $\mathbb{Z}^{(1)} \prec \mathbb{Z}^{(2)}$ and $k_1 = k_2$, $l_1 = l_2$, $m_1 = m_2$. Then*

$$\rho_{AC|\mathbb{Z}^{(1)}}^2 \leq \rho_{AC|\mathbb{Z}^{(2)}}^2$$

Proof. The proof is similar to the proof of Theorem 2 only in this case we can reduce the graphs in each term of the first product to a graph which is Markov equivalent to the graph in 1.i. The terms in the third

product reduce to the DAG in figure 1.ii and those of the second one reduce to figure 3 or figure 2.i, depending on whether one or both of ${}_A\pi_{b_i^{(1)}}$ and ${}_C\pi_{b_i^{(1)}}$ has a collider in $x_{b_i^{(1)}}^* = x_{b_i^{(2)}}^*$. So by the lemmas in section 3 the theorem follows. \square

Theorem 4. *Let $G = (V, E)$ be a singly connected DAG. $A, C \in V$ and $A \neq C$. Suppose ${}_A\pi_C$ has more than one collider. Let $\mathbb{Z}^{(1)} \prec \mathbb{Z}^{(2)}$ and $k_1 = k_2, l_1 = l_2, m_1 = m_2$. Then*

$$\rho_{AC|Z^{(1)}}^2 \leq \rho_{AC|Z^{(2)}}^2$$

Proof. Note that between any two colliders there is exactly one source and conditioning on the nodes on one side of the collider does not change the graph on the other side. Thus one can show that each term in (16) may be reduced to a turn case in which ${}_A\pi_C$ has exactly 1 collider. The proof follows from theorem 3. \square

However if $k_1 \neq k_2, l_1 \neq l_2, m_1 \neq m_2$. Then by the definition of the ordering and the structure of G we note that $k_1 > k_2, l_1 < l_2, m_1 < m_2$. Suppose for instance $b^{(1)} \in N(\mathbb{Z}^{(1)})$ is such that $\exists b_1^{(2)}, b_2^{(2)} \in N(\mathbb{Z}^{(2)})$ s.t. $b^{(1)} \in {}_A\pi_{b_1^{(2)}} \cap {}_C\pi_{b_1^{(2)}} \cap {}_A\pi_{b_2^{(2)}} \cap {}_C\pi_{b_2^{(2)}}$. Then we note that as $\{A, C\} \perp\!\!\!\perp b_2^{(2)} | b^{(1)}, \rho_{AC|\{b^{(1)}, b_2^{(2)}\}}^2 = \rho_{AC|b^{(1)}}^2$.

Thus we redefine $N(\mathbb{Z}^{(1)})$ to be $N(\mathbb{Z}^{(1)}) \cup b_2^{(2)}$ and reduce this case to the case where $k_1 = k_2$. A similar proof follows when $k_2 - k_1 > 1, l_1 < l_2$ and $m_1 < m_2$.

7 Multiply connected DAG

If the graph is not singly connected the ordering of the dependence as described in section 6 may fail. For example consider the graph in figure 6. In figure 7 we plot the squared conditional correlations $\rho_{AC|Y}^2$ and $\rho_{AC|Z}^2$ with b_5 . Other parameter values are fixed at 1. We note that for $b_5 = 0$ $\rho_{AC|Z}^2 < \rho_{AC|Y}^2$ which is a special case for Lemma 3.2. However for smaller and larger values of b_5 the opposite inequality is valid. The range of b_5 in the plot is $[-4, 4]$.

8 Discussion

In this paper we have wrung qualitative information about the strength of dependence from the structure of the graph. We believe that it may be possible to strengthen this result by constructing a richer ordering under which more sets would be comparable.

It is also natural to ask whether the result may be extended to other distributions or classes of graphs. [3]

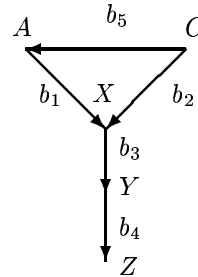


Figure 6: A multiply connected (See Section 7)

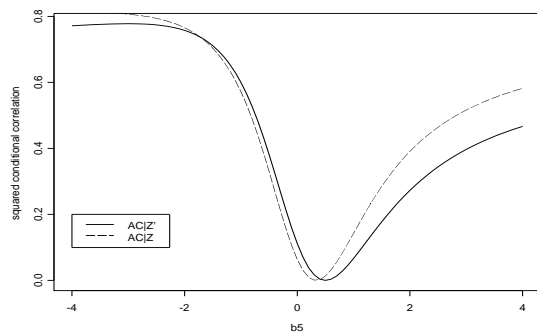


Figure 7: Squared conditional correlation coefficient with b_5 for the DAG in figure 6.

show that discrete distributions exist corresponding to Fig.1.ii in which $I(A;C | Z) > I(A;C | Z')$, hence Lemma 3.2 will not hold without additional assumptions. Section 7 shows that non-monotonic dependence orderings may exist in simple multiply connected Gaussian DAGs. However, in spite of this we believe that the relationships shown in Figure 4 suggest the possibility of developing a theory which provides upper bounds on the strength of conditional dependence which may be present in a given graph.

9 Appendix

By Σ_{BB}^{-1} we denote $(\Sigma_{BB})^{-1}$ and write $x \propto^+ y$ if \exists constant $K > 0$ s.t. $x = Ky$.

Proof of lemma 3.1: Note that

$$\rho_{AC|Z}^2 = \frac{\sigma_{AC|Z}^2}{\sigma_{AA|Z}\sigma_{CC|Z}} \quad (17)$$

$$= \frac{[\sigma_{AC} - \frac{\sigma_{AZ}\sigma_{CZ}}{\sigma_{ZZ}}]^2}{[\sigma_{AA} - \frac{\sigma_{AZ}^2}{\sigma_{ZZ}}][\sigma_{CC} - \frac{\sigma_{CZ}^2}{\sigma_{ZZ}}]} \quad (18)$$

$$\begin{aligned}
&= \frac{[\sigma_{AC}\sigma_{ZZ} - \sigma_{AZ}\sigma_{CZ}]^2}{[\sigma_{AA}\sigma_{ZZ} - \sigma_{AZ}^2][\sigma_{CC}\sigma_{ZZ} - \sigma_{CZ}^2]} \\
&= \frac{[\sigma_{AC}(b_4^2\sigma_{Z'Z'} + \tau_Z^2) - b_4^2\sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{AA}(b_4^2\sigma_{Z'Z'} + \tau_Z^2) - b_4^2\sigma_{AZ'}^2]} \quad (19) \\
&\quad \times \frac{[\sigma_{AC}(b_4^2\sigma_{Z'Z'} + \tau_Z^2) - b_4^2\sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{CC}(b_4^2\sigma_{Z'Z'} + \tau_Z^2) - b_4^2\sigma_{CZ'}^2]}
\end{aligned}$$

$$\begin{aligned}
&= \frac{[a\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]^2}{[a\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2][a\sigma_{CC}\sigma_{Z'Z'} - \sigma_{CZ'}^2]} \quad (20) \\
&\quad (\text{where } a = 1 + \frac{\tau_Z^2}{b_4^2\sigma_{Z'Z'}} \geq 1).
\end{aligned}$$

We shall consider in turn the two terms in this product and compare them to the corresponding terms in $\rho_{AC|Z'}^2$. Consider first

$$\begin{aligned}
&\frac{[a\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[a\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2]} \quad (21) \\
&\quad - \frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2]}.
\end{aligned}$$

After combining, the denominator is always positive so the sign of the expression is determined by the numerator which after some simplification is

$$\begin{aligned}
&(a-1)\sigma_{AZ'}\sigma_{Z'Z'}(\sigma_{Z'C}\sigma_{AA} - \sigma_{AZ'}\sigma_{AC}) \quad (22) \\
&= (a-1)\sigma_{Z'Z'}b_1b_3\tau_X^2(b_2b_3\tau_X^2\sigma_{AA} - b_1^2b_2b_3\tau_X^4) \quad (23)
\end{aligned}$$

$$= (a-1)\sigma_{Z'Z'}b_1b_2b_3^2\tau_X^4(\sigma_{AA} - b_1^2\tau_X^2) \quad (24)$$

$$= (a-1)\sigma_{Z'Z'}b_1b_2b_3^2\tau_X^4(b_1^2\tau_X^2 + \tau_A^2 - b_1^2\tau_X^2) \quad (25)$$

$$= (a-1)\sigma_{Z'Z'}\tau_X^4\tau_A^2b_3^2b_1b_2. \quad (26)$$

Since $a \geq 1$, the above is negative iff $b_1b_2 < 0$. Similarly one can show that.

$$\begin{aligned}
&\frac{[a\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[a\sigma_{CC}\sigma_{Z'Z'} - \sigma_{CZ'}^2]} \quad (27) \\
&\quad - \frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{CC}\sigma_{Z'Z'} - \sigma_{CZ'}^2]}
\end{aligned}$$

$$\alpha^+ (a-1)\sigma_{Z'Z'}\tau_X^4\tau_C^2b_3^2b_1b_2. \quad (28)$$

If $b_1b_2 \geq 0$ then

$$\begin{aligned}
\rho_{AC|Z}^2 &\geq \frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]^2}{[\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2][\sigma_{CC}\sigma_{Z'Z'} - \sigma_{CZ'}^2]} \\
&= \rho_{AC|Z'}^2. \quad (29)
\end{aligned}$$

However

$$\begin{aligned}
&\frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2]} \alpha^+ \sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{Z'C} \\
&= b_1b_2\tau_X^2(\sigma_{Z'Z'} - b_3^2\tau_X^2) \\
&= b_1b_2\tau_X^2\tau_Z^2. \quad (30)
\end{aligned}$$

If $b_1b_2 < 0$

$$\begin{aligned}
&\frac{[a\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[a\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2]} \\
&< \frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{AA}\sigma_{Z'Z'} - \sigma_{AZ'}^2]} < 0 \quad (31)
\end{aligned}$$

$$\begin{aligned}
&\frac{[a\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[a\sigma_{AA}\sigma_{Z'Z'} - \sigma_{CZ'}^2]} \\
&< \frac{[\sigma_{AC}\sigma_{Z'Z'} - \sigma_{AZ'}\sigma_{CZ'}]}{[\sigma_{CC}\sigma_{Z'Z'} - \sigma_{CZ'}^2]} < 0. \quad (32)
\end{aligned}$$

So again

$$\rho_{AC|Z}^2 \leq \rho_{AC|Z'}^2. \quad \square$$

Let K and K' be constants and for some $A, C, D \in V$ and $B \subset V$, where B may be empty, define :

$$L(a) = \frac{[(a-K')\rho_{AC|B} - K\rho_{AD|B}\rho_{CD|B}]^2}{[(a-K') - K\rho_{AD|B}][a - K' - K\rho_{CD|B}]} \quad (33)$$

Then the sign of $\frac{dL(a)}{da}$ is the sign of the numerator of $\frac{dL(a)}{da}$. By the quotient rule

$$\begin{aligned}
&\frac{dL(a)}{da} \alpha^+ 2\rho_{AC|B}[(a-K')\rho_{AC|B} - K\rho_{AD|B}\rho_{CD|B}] \\
&\quad \times [(a-K') - K\rho_{AD|B}][a - K' - K\rho_{CD|B}] \\
&\quad - [a - K' - K\rho_{CD|B}]^2 \\
&\quad \times \{[(a-K') - K\rho_{AD|B}] + [(a-K') - K\rho_{CD|B}]\}. \quad (34)
\end{aligned}$$

Some algebraic simplification yields

$$\frac{dL(a)}{da} \alpha^+ K[(a-K')\rho_{AC|B} - K\rho_{AD|B}\rho_{CD|B}] \quad (35)$$

$$\begin{aligned}
&\times \left\{ [(a-K') - K\rho_{AD|B}]\rho_{CD|B}[\rho_{AD|B} - \rho_{AC|B}\rho_{CD|B}] \right. \\
&\quad \left. + [(a-K') - K\rho_{CD|B}]\rho_{AD|B}[\rho_{CD|B} - \rho_{AC|B}\rho_{AD|B}] \right\}.
\end{aligned}$$

Also note that

$$\rho_{CD|B}[\rho_{AD|B} - \rho_{AC|B}\rho_{CD|B}] \quad (36)$$

$$\begin{aligned}
&\alpha^+ \sigma_{CD|B}[\sigma_{AD|B}\sigma_{CC|B} - \sigma_{AC|B}\sigma_{CD|B}] = M_1, \\
&\rho_{AD|B}[\rho_{CD|B} - \rho_{AC|B}\rho_{AD|B}] \quad (37)
\end{aligned}$$

$$\begin{aligned}
&\alpha^+ \sigma_{AD|B}[\sigma_{CD|B}\sigma_{AA|B} - \sigma_{AC|B}\sigma_{AD|B}] = M_2, \\
&[(a-K')\rho_{AC|B} - K\rho_{AD|B}\rho_{CD|B}] \quad (38)
\end{aligned}$$

$$\alpha^+ [(a-K')\sigma_{AC|B}\sigma_{DD|B} - K \cdot \sigma_{AD|B}\sigma_{CD|B}] = M_3.$$

So

$$\begin{aligned}
&\frac{dL}{da} \alpha^+ KM_3 \left\{ [(a-K') - K\rho_{AD|B}]\sigma_{AA|B}M_1 \right. \\
&\quad \left. + [(a-K') - K\rho_{CD|B}]\sigma_{CC|B}M_2 \right\}. \quad (39)
\end{aligned}$$

In the proofs of Lemmas 3.2, 3.3, 3.4, 3.5 we will find constants K and K' such that

$$K \geq 0, \quad (40)$$

$$(a - K') - K\rho_{AD|B}^2 \geq 0, \quad (41)$$

$$(a - K') - K\rho_{CD|B}^2 \geq 0. \quad (42)$$

In the proof below we shall consider

$$\rho_{AC|BZ}^2 = \frac{[\rho_{AC|B} - \rho_{AZ|B}\rho_{CZ|B}]^2}{[1 - \rho_{AZ|B}^2][1 - \rho_{CZ|B}^2]} \quad (43)$$

We shall then re-express (43) in the form of $L(a)$ and note that $L(1) = \rho_{AC|BZ}'^2$. By using the equation (39) we shall determine the sign of $\frac{dL}{da}$, thereby drawing conclusions on the values of the squared conditional correlations. For lack of space however we present the proof of lemma 3.4 only. The proofs of lemma 3.2 and lemma 3.3 and Lemma 3.5 follow mutatis mutandis.

proof of Lemma 3.4: Let $B = \{B_1, B_2\}$. Then for $B_i \in B$

$$\sigma_{B_i Z} = \frac{\sigma_{ZZ}}{b_6 \tau_{Z'}^2} \sigma_{B_i Z'}. \quad (44)$$

$\sigma_{AZ} = \sigma_{AZ'} = 0$. Also for $Y = \{A, C\}$

$$\sigma_{YZ|B} = \frac{\sigma_{ZZ}}{b_6 \tau_{Z'}^2} \sigma_{YZ'|B}, \quad (45)$$

$$\sigma_{ZZ|B} = \sigma_{ZZ} - \Sigma_{ZB} \Sigma_{BB}^{-1} \Sigma_{BZ} \quad (46)$$

$$= \frac{\sigma_{ZZ}^2}{b_6^2 \tau_{Z'}^4} \left(\frac{b_6^2 \tau_{Z'}^2}{\sigma_{ZZ}} \tau_{Z'}^2 - \Sigma_{Z'B} \Sigma_{BB}^{-1} \Sigma_{BZ'} \right)$$

$$\sigma_{Z'Z'|B} = \tau_{Z'}^2 - \Sigma_{Z'B} \Sigma_{BB}^{-1} \Sigma_{BZ'}. \quad (47)$$

Then with $a = 1/(1 + \frac{\tau_{Z'}^2}{b_6^2 \tau_{Z'}^2}) \leq 1$, $K' = \Sigma_{Z'B} \Sigma_{BB}^{-1} \Sigma_{BZ'} / \tau_{Z'}^2$ and $K = 1 - \Sigma_{Z'B} \Sigma_{BB}^{-1} \Sigma_{BZ'} / \tau_{Z'}^2$, we get

$$\frac{\sigma_{AZ|B} \sigma_{CZ|B}}{\sigma_{ZZ|B}} = \frac{K}{(a - K')} \frac{\sigma_{AZ'|B} \sigma_{CZ'|B}}{\sigma_{Z'Z'|B}}. \quad (48)$$

Using (7) and (43) we can show that $\rho_{AC|BZ}^2$ has the same form as $L(a)$ as in (33). As before $L(1) = \rho_{AC|BZ}'^2$. Then by defining

$$\tilde{\mathbf{b}} = (b_3^{(1)}, b_3^{(2)} b_4) \quad (49)$$

$$\mathbf{b}^* = (b_3^{(1)} b_4 \sigma_{X^{(1)}X^{(1)}}, b_3^{(2)} \sigma_{X^{(2)}X^{(2)}}) \quad (50)$$

$$Q_1 = \tilde{\mathbf{b}} \Sigma_{BB}^{-1} \mathbf{b}^{*T} \quad (51)$$

$$Q_2 = \tilde{\mathbf{b}} \Sigma_{BB}^{-1} \tilde{\mathbf{b}}^T \quad (52)$$

$$Q_3 = b_4 - Q_1 \quad (53)$$

we get

$$\sigma_{AC|B} = \sigma_{AC} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BC} \quad (54)$$

$$= b_1 b_2 b_4 \tau_A^2 - \left(b_1 b_3^{(1)} \tau_A^2, b_1 b_3^{(2)} b_4 \tau_A^2 \right) \Sigma_{BB}^{-1} \begin{pmatrix} b_2 b_3^{(1)} b_4 \sigma_{X^{(1)}}^2 \\ b_2 b_3^{(2)} \sigma_{X^{(2)}}^2 \end{pmatrix}$$

$$= b_1 b_2 \tau_A^2 Q_3 \quad (55)$$

where $\sigma_{X^{(1)}}^2 = \sigma_{X^{(1)}X^{(1)}}$ and similarly for $\sigma_{X^{(2)}}^2$. Similarly we can show

$$\sigma_{CZ'|B} = b_2 b_5 b_6 \tau_{Z'}^2 Q_3 \quad (56)$$

$$\sigma_{AZ'|B} = -b_1 b_5 b_6 \tau_A^2 \tau_{Z'}^2 Q_2. \quad (57)$$

Thus

$$M_1 = -b_1 b_2 b_5^2 b_6^2 \tau_A^4 \tau_{Z'}^4 Q_3 [Q_2 \sigma_{CC|B} + b_2^2 Q_3^2],$$

$$M_2 = -b_1 b_2 b_5^2 b_6^2 \tau_A^4 \tau_{Z'}^4 Q_3 Q_2 [\sigma_{AA|B} + \tau_A^4 Q_2],$$

$$M_3 = b_1 b_2 \tau_A^2 Q_3 [(a - K') \sigma_{Z'Z'|B} + K b_5^2 b_6^2 \tau_{Z'}^2 Q_2].$$

As Q_2 , K and $(a - K')$ are positive the sign of $\frac{dL}{da}$ depends on the sign of $-b_1^2 b_2^2 Q_3^2$, which is negative. So $L(1) \leq L(a)$. Thus it follows that :

$$\rho_{AC|BZ}^2 \geq \rho_{AC|BZ}'^2. \quad \square$$

References

- [1] C. Brito and J. Pearl. Generalised instrumental variables. *UAI 2002*, pages 85–93, 2002.
- [2] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu. Learning Bayesian networks from data: an information-theory based approach. *Artificial Intelligence*, 197:43–90, 2002.
- [3] D.M. Chickering and C. Meek. Monotone DAG faithfulness: a bad assumption. Technical Report MSR-TR-2003-16, Microsoft Research, 2003.
- [4] D. Geiger and J. Pearl. On the logic of causal models. In L. N. Kanal R.D. Schachter, T. S. Levitt and J. F. Lemmer, editors, *Uncertainty in the Artificial intelligence IV*, pages 136–147. North-Holland, Amsterdam, 1990.
- [5] S. Greenland. Quantifying biases in causal models: classical confounding versus collider-stratification bias. *Epidemiology*, 2003. to appear.
- [6] M. Kendall and A. Stuart. *The Advanced Theory of Statistics, vol. 2, Inference and Relationship*. Macmillan Publishing Co., Inc., 1979.
- [7] C. Meek. Strong completeness and faithfulness in Bayesian networks. *UAI 1995*, pages 403–410.
- [8] T. Richardson and P. Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.
- [9] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search*. Number 81 in Lecture Notes in Statistics. 1993.
- [10] J. Whittaker. *Graphical Models in Applied Mathematical Multivariate Statistics*. John Wiley & Sons, 1989.