DEMONSTRATIO MATHEMATICA
Vol. XXXIII No 1 2000

Mieczysław A. Kłopotek

# ON A DEFICIENCY OF THE FCI ALGORITHM LEARNING BAYESIAN NETWORKS FROM DATA

**Abstract.** Causally insufficient structures (models with latent or hidden variables, or with confounding etc.) of joint probability distributions have been subject of intense study not only in statistics, but also in various AI systems. In AI, belief networks, being representations of joint probability distribution with an underlying directed acyclic graph structure, are paid special attention due to the fact that efficient reasoning (uncertainty propagation) methods have been developed for belief network structures. Algorithms have been therefore developed to acquire the belief network structure from data. As artifacts due to variable hiding negatively influence the performance of derived belief networks, models with latent variables have been studied and several algorithms for learning belief network structure under causal insufficiency have also been developed. Regrettably, some of them are known already to be erroneous (e.g. IC algorithm of [12]). This paper is devoted to another algorithm, the Fast Causal Inference (FCI) Algorithm of [17]. It is proven by a specially constructed example that this algorithm, as it stands in [17], is also erroneous. Fundamental reason for failure of this algorithm is the temporary introduction of non-real links between nodes of the network with the intention of later removal. While for trivial dependency structures these non-real links may be actually removed, this may not be the case for complex ones, e.g. for the case described in this paper. A remedy of this failure is proposed.

## 1. Introduction

Various expert systems, dealing with uncertain data and knowledge, possess knowledge representation in terms of a belief network (e.g. knowledge base of the MUNIM system [1] , ALARM network [2] etc.). A number of efficient algorithms for propagation of uncertainty within belief networks and their derivatives have been developed, compare e.g. [11, 13, 14].

Belief networks, causal networks, or influence diagrams, are terms frequently used interchangeably. They are quite popular for expressing causal relations under multiple variable setting both for deterministic and non-deterministic (e.g. stochastic) relationships in various domains: statistics,

**Keywords:** Belief networks, discovery under causal insufficiency,

philosophy, artificial intelligence [3, 16]. Though a belief network (a representation of the joint probability distribution, see [3]) and a causal network (a representation of causal relationships [16]) are intended to mean different things, they are closely related. Both assume an underlying dag (directed acyclic graph) structure of relations among variables and if Markov condition and faithfulness condition [17] are met, then a causal network is in fact a belief network. The difference comes to appearance when we recover belief network and causal network structure from data. A dag of a belief network is satisfactory if the generated probability distribution fits the data, may be some sort of minimality is required. A causal network structure may be impossible to recover completely from data as not all directions of causal links may be uniquely determined [17]. Fortunately, if we deal with causally sufficient sets of variables (that is whenever significant influence variables are not omitted from observation), then there exists the possibility to identify the family of belief networks a causal network belongs to [18] (see also [7]).

Regrettably, to our knowledge, a similar result is not directly known for causally insufficient sets of variables (that is when significant influence variables are hidden) - "Statistical indistinguishability is less well understood when graphs can contain variables representing unmeasured common causes" ([17], p. 88). Latent (hidden) variable identification has been investigated intensely both for belief networks (e.g. [10, 6, 9, 2]) and causal networks ([12, 16, 17, 4, 5]), beside the immense research effort in traditional statistics (to mention results of Spearman on vanishing tetrad differences from the beginning of this century to recent LISREL and EQS techniques - see [15] for a comparative study of these techniques with causal network approaches in AI). The algorithm of [2] recovers the most probable location of a hidden variable. Whereas the CI algorithm of [17] recovers exact locations of common causes, but clearly not all of them. In fact, the CI algorithm does not provide a dag, but rather a graph with edges fully (unidirected or bidirected) or partially oriented, or totally non-oriented with additional constraints for edge directions at other edges. Partially or non-oriented edges may prove to be either directed or bidirected edges. Alternatively, the IC algorithm of Pearl and Verma [12] tried to recover the family of "minimal latent models" (a family of dags close to the data), but, as Spirtes et al. claim in [17], page 200, "Unfortunately, the two main claims about the output of the Inductive Causation Algorithm made in the paper ... are false.". Hence the big question is whether or not the bidirectional edges (that is indications of a common cause) are the only ones necessary to develop a belief network out of the product of CI, or must there be some other hidden variables added (e.g. by guessing). We answer this question in favour of the CI algorithm elsewhere [8].

However, as Spirtes et al. state, their CI algorithm is feasible only for a small number of variables and hence they developed an "accelerated" version of the CI algorithm: the FCI algorithm, which also has a partial including path graph as its output. The question formulated for CI needs thus to be repeated for the FCI algorithm. Regrettably as it is, the FCI algorithm, as it stands in [17], cannot be accommodated for recovery of possible belief networks as it introduces into the causal structure causal arrows which are not actually present in the data, and due to this fact a resulting belief network would contain dependencies not present in the data, but what is worse, it would exhibit independencies not present in the data.

We are indebted to an anonymous Referee for a hint on the intended meaning of the FCI algorithm and based upon this hint we present here a corrected version of FCI.

## 2. Fast causal inference algorithm of Spirtes et al.

To make this paper self-contained, we below remind the Causal Inference (CI) and the Fast Causal Inference (FCI) algorithms of Spirtes, Glymour and Scheines [17] together with some basic notation used therein. Recalling CI algorithm is necessary as FCI refers to CI in its final phase. The text of this section is to a large extent a citation from [17], and quotation marks will be dropped for readability.

Essentially, the CI algorithm recovers partially the structure of an including path graph. Given a directed acyclic graph G with the set of hidden nodes $V_h$ and visible nodes $V_s$ representing a causal network CN, an including path between nodes A and B belonging to $V_s$ is a path in the graph G such that the only visible nodes (except for A and B) on the path are those where edges of the path meet head-to-head and there exists a directed path in G from such a node to either A or B. An including path graph for G is such a graph over $V_s$ in which if nodes A and B are connected by an including path in G ingoing into A and B, then A and B are connected by a bidirectional edge $A < - > B$. Otherwise if they are connected by an including path in G outgoing from A and ingoing into B then A and B are connected by an unidirectional edge $A- > B$.

A partially oriented including path graph contains the following types of edges unidirectional: $A- > B$, bidirectional $A < - > B$, partially oriented $Ao- > B$ and non-oriented $Ao - oB$, as well as some local constraint information $A * -\underline{*B*} - *C$ meaning that edges between A and B and between B and C cannot meet head to head at B. (Subsequently an asterisk ($*$) means any orientation of an edge end: e.g. $A * - > B$ means either $A- > B$ or $Ao- > B$ or $A < - > B$).

In a partially oriented including graph $\pi$ (see [17], pp.: 181-182) we have

(i) A is a parent of B if and only if $A-> B$ in $\pi$.

(ii) B is a collider along the path $< A, B, C >$ if and only if $A *-> B <-* C$ in $\pi$. B is a definite non-collider on undirected path U if and only if either B is an end-point of U, or there exist vertices A and C such that U contains one of the subpaths $A <-- B *-* C$, $A *-* B --> C$, or $A * -\underline{*B*} - *C$, (see Glossary of [17]).

(iii) An edge between B and A is into A iff $A <-* B$ in $\pi$

(iv) An edge between B and A is out of A iff $A-> B$ in $\pi$.

(v) A is d-separated from B given set S iff A and B are conditionally independent given S.

(vi) A and B are d-connected given node C iff there exists no such set S containing C such that A and B are conditionally independent given S.

(vii) In a partially oriented including path graph $\pi'$, U is a definite discriminating path for B if and only if U is an undirected path between X and Y containing B, $B \neq X, B \neq Y$, every vertex on U except for B and the endpoints is a collider or a definite non-collider on U and:
(a) if V and V" are adjacent on U, and V" is between V and B on U, then $V *-> V"$ on U,
(b) if V is between X and B on U and V is a collider on U, then $V-> Y$ in $\pi$, else $V <-* Y$ on $\pi$
(c) if V is between Y and B on U and V is a collider on U, then $V-> X$ in $\pi$, else $V <-* X$ on $\pi$
(d) X and Y are not adjacent in $\pi$.

viii) U is a directed path from X to Y iff there exists an undirected path between X and Y such that if V is adjacent to X on U then $X-> V$ in $\pi$, if $V$ is adjacent to Y on V, then $V-> Y$, if V and V" are adjacent on U and V is between X and V" on U, then $V-> V"$ in $\pi$.

**The Causal Inference (CI) Algorithm:** (see [17], pp.: 183)
Input: Empirical joint probability distribution
Output: partial including path graph $\pi$.

**A)** Form the complete undirected graph Q on the vertex set V.

**B)** if A and B are d-separated given any subset S of V, remove the edge between A and B, and record S in Sepset(A,B) and Sepset(B,A).

**C)** Let F be the graph resulting from step B). Orient each edge o-o. For each triple of vertices A,B,C such that the pair A,B and the pair B,C are each adjacent in F, but the pair A,C are not adjacent in F, orient $A *-* B *-* C$ as $A *-> B <-* C$ if and only if B is not in Sepset(A,C), and orient $A *-* B *-* C$ as $A * -\underline{*B*} - *C$ if and only if B is in Sepset(A,C).

**D)** Repeat

if there is a directed path from A to B, and an edge $A*-*B$, orient $A*-*B$ as $A*->B$,

else if B is a collider along $< A, B, C >$ in $\pi$, B is adjacent to D, and A and C are not d-connected given D, then orient $B*-*D$ as $B<-*D$ ,

else if U is a definite discriminating path between A and B for M in $\pi$ and P and R are adjacent to M on U, and P-M-R is a triangle, then

if M is in Sepset(A,B) then M is marked as non-collider on subpath $P*-\underline{*M*}-R$

else $P*-*AM*-*R$ is oriented as $P*->M<-*R$,

else if $P*-\underline{>M*}-*R$ then orient as $P*->M->R$.

until no more edges can be oriented.

**This ends the algorithm CI.**

To understand the proper FCI algorithm, some additional definitions are necessary:

ix) In a full including graph $\pi_0$ V is in D-Sep(A,B) iff $V \neq A$ and there is an undirected path from V to A such that all the nodes on the path are colliders having either A or B as their definite successor. (see [17], p. 187)

x) "For a given partially constructed partially oriented including path graph $\pi$, **Possible-D-Sep**(A,B) is defined as follows: If $A \neq B$, V is in **Possible-D-Sep**(A,B) in $\pi$ if and only if $V \neq A$, and there is an undirected path U between A and V in $\pi$ such that for every subpath $< X, Y, Z >$ of U either Y is a collider on the subpath, or Y is not a definite non-collider and on U, and X, Y, and Z form a triangle in $\pi$. " ([17], p.187 below Fig.18, repeated in Glossary therein).

**The Fast Causal Inference (FCI) Algorithm:** (see [17], p.: 188)
Input: Empirical joint distribution
Output: partial including path graph $\pi$.

**A)** Form the complete undirected graph Q on the vertex set V.

**B)** n = 0;

repeat

repeat

select an ordered pair of variables X and Y that are adjacent in Q such that Adjacencies(Q,X)-{Y} has cardinality greater or equal to n, and a subset S of Adjacencies(Q,X)-{Y} of cardinality n, and if X and Y are

independent given S delete edge between X and Y from Q, and record S in Sepset(X,Y) and in SepSet(Y,X)
until all ordered pairs of adjacent variables such that Adjacencies(Q,X) -{Y} has cardinality greater than or equal to n and all subsets S of Adjacencies(Q,X)-{Y} of cardinality n have been tested for making X,Y independent.;
n=n+1;
until for each ordered pair of adjacent vertices X,Y, Adjacencies(Q,X)-{Y} is of cardinality less than n.

**C)** Let F be the undirected graph resulting from step B). Orient each edge as $o - o$. For each triple of vertices A,B,C such that the pair A,B and the pair B,C are each adjacent in F, but the pair A,C are not adjacent in F, orient $A * - * B * - * C$ as $A * - > B < - * C$ if and only if B is not in Sepset(A,C).

**D)** For each pair of variables A and B adjacent in F', if A and B are independent given any subset S of Possible-D-SEP(A,B)-{A,B} or any subset S of Possible-D-SEP(B,A)-{A,B} in F remove edge between A and B and record S in Sepset(A,B) and Sepset(B,A).

**E)** Reset all edge orientations as $o - o$ and carry out steps C) and D) of the Ci algorithm

**This ends the algorithm FCI**

## 3. The claim of this paper about FCI

Let us imagine that we have obtained a partial including path graph from FCI, and we want to find a Belief Network representing the joint probability distribution out of it. Let us consider the following algorithm:

**FCI-to-BN Algorithm**

Input: Result of the FCI algorithm (a partial including path graph)
Output: A belief network

**A)** Accept unidirectional and bidirectional edges obtained from CI.

**B)** Orient every edge $Ao- > B$ as $A- > B$.

**C)** Orient edges of type $Ao - oB$ either as $A < -B$ or $A- > B$ so as not to violate $P * - \underline{* M *} - * R$ constraints.

**This ends the algorithm CI-to-BN**

We claim that:

THEOREM. The belief network obtained via FCI-to-BN algorithm does not in general keep all the dependencies and independencies of the original underlying including path graph.

The rest of this paper provides a sketchy proof of the above theorem by an example. First we demonstrate, that step C) of FCI generates arrow orientations contradicting the edge orientation of the original including path graph. Then we show that this leads to violation of dependence/independence relation in the resulting belief network. We then suspect that definition (x) above of Possible-D-SEP is not correct and check another meaning thereof. Though it provides a recovery from the failure of the initial example, it runs into error on a larger example. Finally, following a hint from an anonymous Referee we rewrite FCI algorithm altogether to ensure in step D removal of superfluous edges left in step B.

## 4. FCI, as It stands, fails

Please compare first definitions (ii) of definite non-collider and (x) of Possible-D-Sep with the contents of proper FCI algorithm. Node Y from definition (x) is not the endpoint on U, proper FCI algorithm introduces neither unidirectional edges nor $X*-\underline{*Y*}-*Z$ constraints. Hence the phrase "Y is not a definite non-collider" in definition (x) is absolutely pointless as Y is always not a definite non-collider out of the construction of FCI algorithm, as it stands in [17]. We rewrite definition (x) as:

x') For a given partially constructed partially oriented including path graph $\pi$, **Possible-D-Sep**(A,B) is defined as follows: If $A \neq B$, V is in **Possible-D-Sep**(A,B) in $\pi$ if and only if $V \neq A$, and there is an undirected path U between A and V in $\pi$ such that for every subpath $< X, Y, Z >$ of U either Y is a collider on the subpath, or Y is on U, and X, Y, and Z form a triangle in $\pi$.

Let us study a run of the FCI algorithm on a set of visible (observable) variables with intrinsic causal relationships from Fig. 1. The double arrows $A < - > B$ in this figure are to be interpreted as follows: there exists a (hidden, not observable) variable $H_{A,B}$ such that the causal relationship is in fact as follows: $A < -H_{A,B}- > B$.

Step A) of FCI is trivial. Let us consider step B). We start with n = 0.

We obtain the undirected graph in Fig. 2. Please notice at this stage, that there are three edges $Y_3 - X_3$, $S_3 - W_3$ and $R_3 - V_3$ not present in the original graph of Fig. 1. We shall not be alerted by this fact as the step D of FCI possibly removes further edges.

Let us turn to step C of FCI. We orient stepwise edges - see Fig. 3.

This is in agreement with the original graph up to the following edges: $Y_3 < - > X_3$, $S_3 < - > W_3$ and $R_3 < - > V_3$ which are superfluous and $P_3 < - > X_3$ oriented contradictory to intention of the original graph:

In this way we obtain the partial including path graph of Fig. 3.

We arrive at step D) of the algorithm.

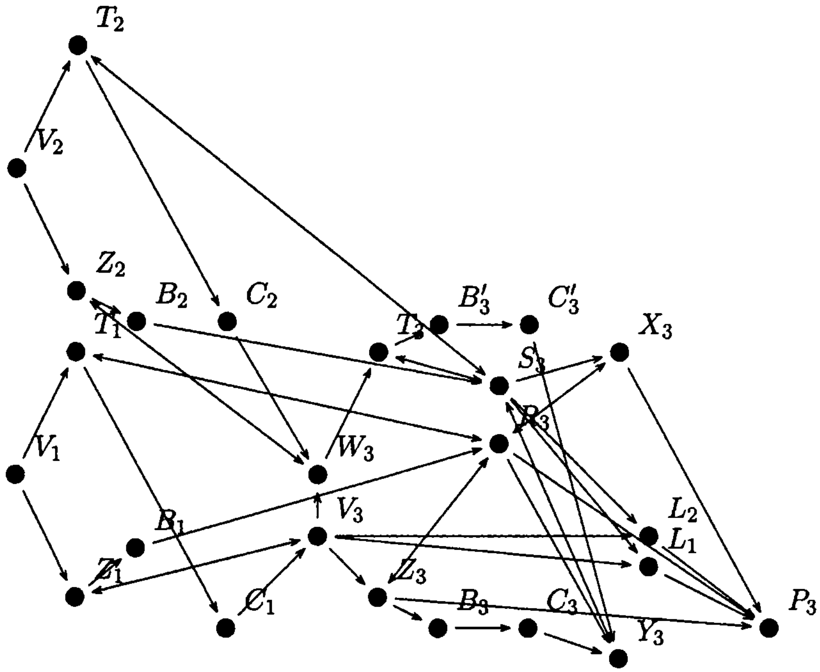


Figure 1: Original network

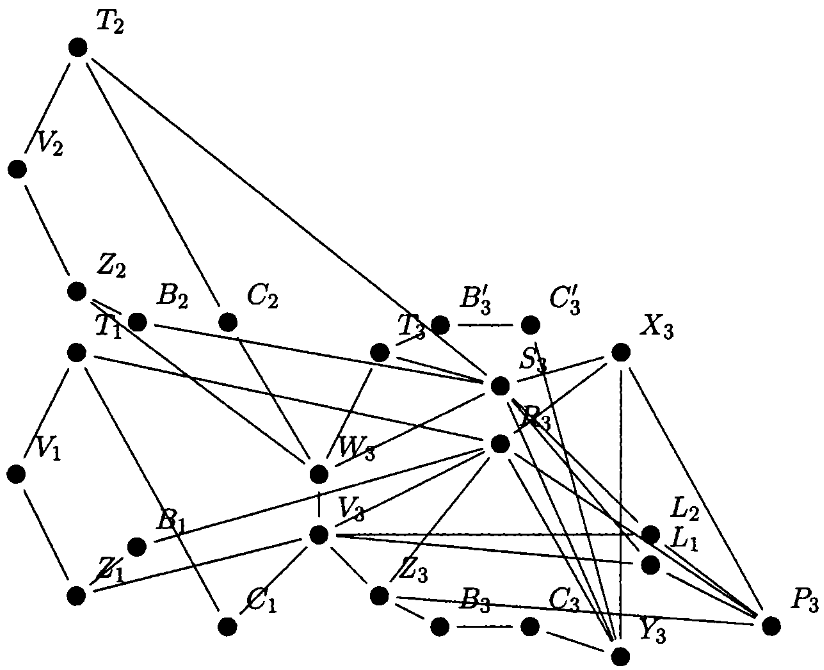


Figure 2: After FCI Algorithm step B

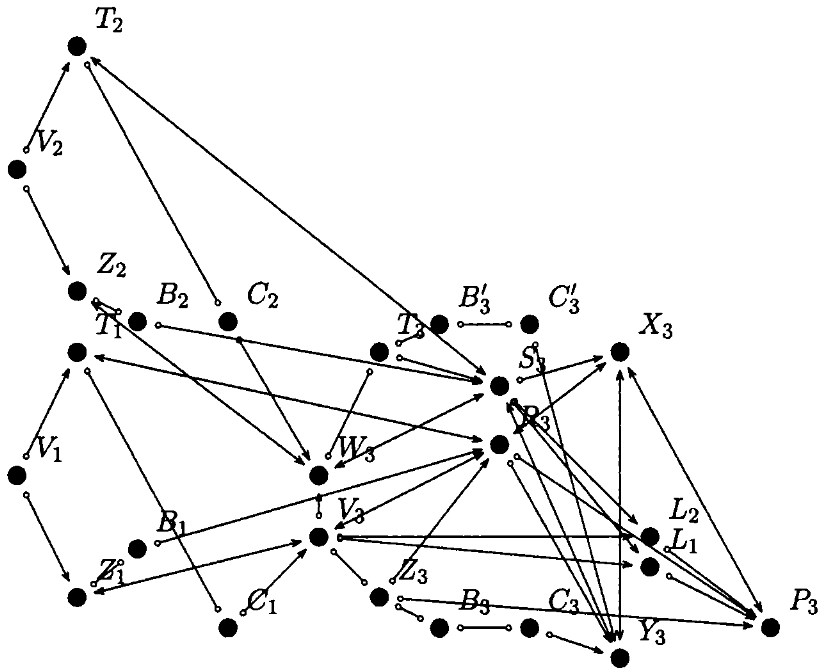


Figure 3: After FCI Algorithm step C

```
FCI stage D protocol (-only a part thereof)
Original network:
    Between nodes R3 and V3  D-Sep {T1,V1,X3,Z3,b1}
FCI-Derived network:
    Between nodes R3 and V3
(possible) D-Sep {P3,L1,L2,X3,S3,Y3,C3,c3,T2,V2,T3,W3,Z2,c2,b2,Z3,T1,V1,b1}
Edge removal:  ~eR3   V3 SepSet {T1,V1,b1}

Original network:
    Between nodes S3 and W3 D-Sep {T2,V2,T3,Y3,b2}
FCI-Derived network:
    Between nodes S3 and W3
(possible) D-Sep {L1,V3,L2,T2,V2,T3,X3,P3,R3,Z3,T1,V1,b1,Y3,C3,c3,b2}
Edge removal:  ~eS3   W3 SepSet {T2,V2,b2}

Original network:
    Between nodes X3 and Y3 D-Sep {R3,T1,V1,Z3,V3,Z1,c1,b1,S3}
FCI-Derived network:
    Between nodes X3 and Y3 (possible) D-Sep {P3,L1,L2,R3,Z3,T1,V1,b1,S3}
```

```
Original network:
    Between nodes Y3 and X3 D-Sep {C3,R3,S3,T2,V2,T3,W3,b2,c3}
FCI-Derived network:
    Between nodes Y3 and X3 (possible) D-Sep {C3,R3,S3,T2,V2,T3,b2,c3}
```

Two of the unwanted edges $S_3 <-> W_3$ and $R_3 <-> V_3$ are removed correctly, but the third $Y_3 <-> X_3$ not due to nodes of D-Sep missing in Possible-D-Sep.

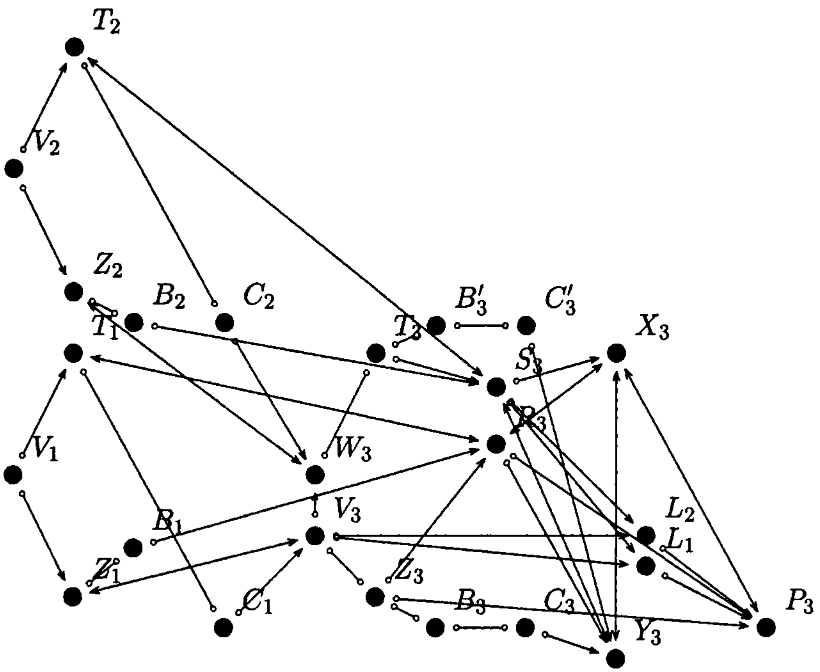


Figure 4: After FCI Algorithm step D

As a result we obtain the graph of Fig. 4. If we apply now step E) of FCI obtaining erroneous edge $Y_3 <-> X_3$ and erroneous edge orientation $P_3 <-> X_3$ indicating erroneous (conditional) dependence of $Y_3$ on $X_3$ (within the original graph e.g. $\{R_3, S_3, Z_3, V_3\}$ d-separates both) and conditional independence of $S_3$ and $P_3$ on $\{L_1, L_2, L_3, L_4\}$ whereas in the original network also $X_3$ is needed to d-separate both. We obtain also a contradictory information: the constraint $S_3 * - \underline{* X_3 *} - P_3$ and at the same time edge orientations: $S_3 -> X_3 <-> P_3$.

Notice that failure to remove edge $Y_3 <-> X_3$ is related to the sequence of checking edges for removal. If this edge were tried first, then no error would occur.

By the above Theorem 1 is proven.

## 5. Repairing FCI algorithm

Posible-D-Sep should always be superset of D-Sep and errors in orientation of intrinsic edges resulting from presence of superfluous edges in a partially including paths graphs as well as presence of superfluous edges themselves should not remove any D-Sep nodes from Possible-D-Sep. In this case, however, both the definition of Possible-D-Sep and the FCI algorithm itself need to be modified. They should not rely on presence of arrows at edge ends because they are a property which is neither truth-preserving nor falsehood-preserving on removal of superfluous edges. The truth-preserving property in that case is the presence of local constraint information $A * -\underline{*B*} - *C$. Therefore let us redefine the notion of Possible-D-Sep as follows:

x"') For a given partially constructed partially oriented including path graph $\pi$, **Possible-D-Sep**(A,B) is defined as follows: If $A \neq B$, V is in **Possible-D-Sep**(A,B) in $\pi$ if and only if $V \neq A$, and there is an undirected path U between A and V in $\pi$ such that for every subpath $< X, Y, Z >$ of U we have *no* local constraint information $X * -\underline{*Y*} - *Z$ in $\pi$.

Furthermore stage C of original FCI algorithm has to be replaced by the following prescription:

**C')** Let F be the undirected graph resulting from step B). Orient each edge as $o - o$. For each triple of vertices A,B,C such that the pair A,B and the pair B,C are each adjacent in F, but the pair A,C are not adjacent in F, orient $A * - * B * - * C$ as $A * -\underline{*B*} - *C$ if and only if B is in Sepset(A,C).

Summarizing, both definition of Possible-D-Sep and the stage C) of FCI have to consume or produce resp. local constraint information instead of head-to-head edge orientation ("colliders"). This in my opinion corrects the algorithm completely and correctly. To prove this claim briefly, let us first turn to relationship between Possible-D-Sep and D-Sep in a fully oriented intrinsic including path graph. Obviously, any node in D-Sep will also belong to Possible-D-Sep as the path out of collider nodes in D-Sep excludes any local constraint information $X * -\underline{*Y*} - *Z$ for any three subsequent nodes on this path. Let us now consider a partially constructed partially oriented including path graph, if $X - Y$ and $Y - Z$ are intrinsic connections then $X * -\underline{*Y*} - *Z$ information in the partially constructed partially including path graph would immediately imply that $X * -\underline{*Y*} - *Z$ holds also in the intrinsic underlying including path graph - as this relies not on graph properties but solely on conditional independence property. Hence nothing like this appears on an intrinsic collider path, hence Possible-D-Sep will always contain D-Sep.

Notice that in a partially constructed partial including path graph superfluous local constraints information $A{*}{-}\underline{{*}B{*}}{-}{*}C$ may occur in case that for example edge $A-B$ is superfluous one. But this does not disturb the algorithm in any way as does not influence relationship between Possible-D-Sep and D-Sep. In some sense it may be considered as a correct information, because an intrinsic edge can never meet head-to-head with a non-existent edge.

## 6. Discussion

The paper demonstrates by example that the FCI algorithm as it stands in [17], is not correct. To repair it, it is necessary either to drop step C) spoiling the whole algorithm altogether, or to change both the definition of Possible-D-Sep and step C. This is because of the philosophy of FCI: it is meant to remove first as much edges as possible using only direct neigbbours of a node (just taking advantage of earlier edge removals in the process), and then to take into account those nodes, which are not neighbouring but influence dependence relationship between the nodes which is done in step D). Step C) was intended as a way to bind set of potential candidates for dependency considerations of step D), but is just demonstrated to be wrong. Obviously, it can be removed without violating the philosophy of the algorithm, while re-establishing algorithm's correctness. FCI algorithm modified by removal of step C), however, would not be too beneficial compared to the primary CI algorithm but for really sparse networks. And hence, like CI, would be rarely applicable to networks larger than a few nodes.

Therefore the alternative presented in previous section seems to be reasonable. It is, however, more space consuming - due to the necessity of maintaining local constraint list. We could instead take the policy - as an alternative to section 6 approach - that we would rerun steps C and D whenever an edge has been removed in stage D. This seems however not to be a time-efficient solution.

Still another alternative could be to postpone removal of an edge detected as superfluous in Stage D until the test D is completed for all the other edges. This approach would require only one logical cell of storage for each edge (to notify whether or not the edge has to be removed after completion of stage D.). However, as with the original FCI algorithm, the orientation information for edges would then have to be dropped and later recalculated in stage C of CI.

Two facts about the sample network used in the proof of Theorem 1 cannot be overseen: the network as such is rather a big one and this network is artificially constructed, and therefore may seldom occur in practice. However, from the point of view of statistics it is not negligible that an algorithm makes systematic errors beside random ones.

The lesson to be learned from failure of FCI is that one should be very careful if a network structure algorithm runs at risk of introducing non-existent links between nodes, especially if it is based on local criteria like FCI and CI.

The result of this paper has consequences for the validity of the theory developed in Chapter 6 of [17]. Our result means directly that Theorem 6.4 of [17] is wrong, and all the claims derived from it should be at least reconsidered.

## 7. Conclusions

In this paper non-suitability of the FCI algorithm of Spirtes et al, as it stands in [17], for recovery of belief networks from data under causal insufficiency has been proven by example. It must be acknowledged that the size of the demonstration network is considerable, and hence under practical settings under which this type of algorithms is applied, should seldom have a chance to emerge. However, users of the algorithm should be aware of possibilities of occasional failures built into the philosophy of the algorithm.

The only ad hoc possibility of repair for FCI is to drop its step C) altogether, but then improvement over CI of [17] is only for very sparse networks, and CI is known to be feasible for networks with mean and large number of nodes.

A more elaborate repair method has been proposed which changes the definition of Possible-D-Sep and stage C of FCI algorithm.

Further research efforts are necessary to establish other derivatives of the CI algorithm which would be computationally feasible but not lead to incorrect results by definition.

### References


[1] S. Andreassen, M. Woldbye, B. Falck, S. K. Andersen, *MUNIN - a causal probabilistic network for interpretation of electromyographic findings*, Proc. of the Tenth International Joint Conference on AI (1987), 366–372.

[2] G. F. Cooper, E. Herskovits, *A Bayesian method for the induction of probabilistic networks from data*, Machine Learning 9 (1992), 309–347.

[3] D. Geiger, T. Verma, J. Pearl, *d-Separation: From Theorems to Algorithms*, M. Henrion, R. D. Shachter, L. N. Kamal, J. F. Lemmer (Eds.): Uncertainty in Artificial Intelligence 5, Elsevier Science Publishers B.V. (North-Holland), (1990), pp. 139–148.

[4] C. Glymour, R. Scheines, P. Spirtes, K. Kelley, *Discovering Causal Structure*, New York, Academic Press, 1987.

[5] C. Glymour, P. Spirtes, *Latent variables, causal models and overidentifying constraints*, J. Econometrics, 39 (1988), 175–198.

[6] J. L. Golmard, A. Mallet, *Learning probabilities in causal trees from incomplete databases*, Proc. of the IJCAI Workshop on Knowledge Discovery in Databases (Detroit M.I.) (1989), 117–126

[7] M. A. Kłopotek, *Partial dependency separation—a new concept for expressing dependence/independence relations in causal networks*, Demonstratio Math. 32 (1999), 207–226.

[8] M. A. Kłopotek, *Learning Belief Network Structure From Data under Causal Insufficiency*, in: F. Bergadano, L. DeRaed (Eds.): Machine Learning ECML-94, Lecture Notes in Artificial Intelligence 784, Springer-Verlag, 1994, pp. 379–382.

[9] L. Liu, D. C. Wilkins, X. Ying, Z. Bian, *Minimum error tree decomposition*, Proc. of the Conference on Uncertainty in Artificial Intelligence, Cambridge M.A. (1990), 54–62.

[10] J. Pearl, *Fusion, propagation and structuring in belief networks*, Artificial Intelligence, 29 (1986), 241–288.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo CA, (1988.

[12] J. Pearl, T. Verma, *A theory of inferred causation*, [in:] Principles of Knowledge Representation and Reasoning, Proc. of the Second International Conference, Cambridge, Massachusetts, April 22-25, Allen J., Fikes R., Sandewell E. Eds„ San Mateo CA:Morgen Kaufmann, (1991), 441–452.

[13] R. D. Shachter, *Evidence absorption and propagation through evidence reversals*, in: M. Henrion, B. D. Shachter, L. N. Kanal, J. F. Lemmer: Uncertainty in Artificial Intelligence 5, Elsevier Science Publishers B.V (North-Holland), (1990), 173–190.

[14] P. P. Shenoy, G. Shafer, *Axioms for probability and belief-function propagation*, in: R. D. Shachter, T. S. Levit, L. N. Kanal, J. F. Lemmer eds: Uncertainty in Artificial Intelligence 4, (Elsevier Science Publishers B.V. (North Holland) (1990).

[15] P. Spirtes, C. Glymour, R. Scheines, *Simulation studies of the reliability of computer-aided model specification using TETRAD II, EQS and LISREL programs*, Sociological Methods and Research, 19 (1990), 3–66.

[16] P. Spirtes, C. Glymour, R. Scheines, *Causality from probability*, in: G. McKee ed.: Evolving knowledge in natural and artificial intelligence (London: Pitman), (1990), 181–199.

[17] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction and Search*, Lecture Notes in Statistics 81, Springer-Verlag, (1993).

[18] T. Verma, J. Pearl, *Equivalence and synthesis of causal models*, Proc. 6th Conference on Uncertainty in AI (1990), 220–227.

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES, WARSAW
ul. Ordona 21
01-237 WARSAW, POLAND
and
FACULTY OF MATHEMATICS AND INFORMATION SCIENCES
WARSAW UNIVERSITY OF TECHNOLOGY
Pl. Politechniki 1
00-661 WARSAW, POLAND
e-mail: klopotek@ipipan.waw.pl, http://www.ipipan.waw.pl/~klopotek/

*Received April 20, 1999.*