Applying Graphical Models to Partially Observed Data Generating Processes

Rebecca Ayesha Ali

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

University of Washington

2002

Program Authorized to Offer Degree: Statistics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Rebecca Ayesha Ali

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

Thomas S. Richardson

Reading Committee:

Michael D. Perlman

Babette A. Brumback

Date:

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to Bell and Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature______________________________

Date______________________________

University of Washington

Abstract

Applying Graphical Models to Partially Observed Data Generating Processes

by Rebecca Ayesha Ali

Chair of Supervisory Committee:

Professor Thomas S. Richardson
Statistics

Model specification is an important part of statistical analyses, particularly in the absence of background knowledge. This thesis makes advances in graphically representing partially observed processes in a way that can facilitate the model search process. Graphical Markov models have been used in fields as diverse as economics, social sciences, computer science, and engineering, particularly when the data-generating process is unknown. Structural equation models from economics, path diagrams from the social sciences, Bayesian networks from computer science and Kalman filters from engineering are examples of graphical models.

Ancestral graphs provide a class of graphs that can encode conditional independence relations that arise in directed acyclic graph (DAG) models with latent and selection variables, corresponding to marginalization and conditioning. However, for any ancestral graph, there may be several other graphs to which it is Markov equivalent. I introduce a simple representation of a Markov equivalence class for ancestral graphs, thereby facilitating the model search process for a model given some data.

More specifically, I define a join operation on ancestral graphs which will associate a unique graph with an equivalence class. I extend the separation criterion for ancestral graphs (which is an extension of d-separation) and provide a proof of the pairwise and global Markov properties for joined ancestral graphs. I also demonstrate that *minimal*

j-connecting paths, (see Definition 3.1.1) take on a special structure.

Keywords: maximal ancestral graphs, joined graphs, Markov equivalence, DAG models, latent and selection variables, d-separation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere thanks to Thomas Richardson for his support and guidance throughout my graduate program. He has taught me so much about teaching, conducting academic research, and career-planning, and I am grateful to have been so fortunate as to have worked with him. I would also like to thank my other committee members: Babette Brumback for introducing me to the interface of graphical models and epidemiology, and for always giving me the new perspective for looking at data; to Michael Perlman for his confidence in me and his valuable comments on both technical and philosophical aspects of my dissertation; and to Alejandro Murua for taking time out to sincerely think about questions related to my work in Electrical Engineering.

I would also like to thank Laura Koutsky and Jim Hughes for their patience and enthusiasm. I have learned much about the practical aspects of doing causal inference from our work together on cervical cancer.

Finally, I would like to thank my family for their unwavering support, even if it meant supporting me to move away from the rest of our family. To my Dad: thank-you for showing me the "beauty in math". Throughout my academic career, you have guided me in all aspects, both statistical and non-academic, of my academic training. I thank my brothers Sadek and Adnan Ali for taking time to help me with programming questions at any time of the day or night. I thank my sister Sharmene Ali for offering me advice on graduate school and letting me learn from her experiences. Thanks to her, my first publication was a family collaboration. And I thank the rest of my brothers and sisters for supporting me, even financially. And to my nephews whose pictures made my time in my office more enjoyable.

The work in this thesis was supported by the U.S. National Science Foundation under grant DMS-9972008, and by the U.S. Environmental Protection Agency.

DEDICATION

I would like to dedicate this thesis to my parents, Suraiya and Mir M. Ali.

Chapter 1

INTRODUCTION

This thesis deals with the problem of model selection for models that can be represented by graphical models with latent variables. Chapter 1 provides background on graphical Markov models and reviews relevant results presented by other researchers. Chapter 2 introduces a new class of graphs: joined graphs. The main results of this thesis are presented in Chapters 2 (Theorem 2.6.1) and 3 (Theorem 3.8.1) in which proofs of maximality and the global Markov property for this new class of graphs are respectively provided. The pairwise Markov property follows from the global Markov property for these graphs. In Chapter 3, Corollary 3.2.1 outlines the special structure of *minimal j-connecting paths* (See Definition 3.1.1). Finally Chapter 4 will summarize what has been presented in the previous chapters and discuss directions for future related research.

Graphical Markov models have been used in fields as diverse as economics, social sciences, computer science, and engineering, particularly when the data-generating process is unknown. Structural equation models, path diagrams, and Bayesian networks are examples of graphical models.

> *Many of the classical multivariate probabilistic systems studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics are special cases of the general graphical model formalism – examples include mixture models, factor analysis, hidden Markov models, Kalman filters and Ising models. The graphical model framework provides a way to view all of these systems as instances of a common underlying formalism.*
> (Michael Jordan, 1998)

1.1 Background of Graphical Markov Models

A graphical Markov model is a set of distributions that can be described by a graph consisting of vertices and edges. In particular, suppose there is some data-generating process such as treatment of a disease: the variables relevant to that system would form the vertices (or nodes) of the graph, and the presence or absence of an edge between any two variables encodes the conditional independence relations holding among the vertices in the graph. The different tasks of graphical models can be divided into three stratum:

(i) If the structure of the graph is known and the parameters associated with the conditional distributions encoded by the graph are known, then one can do inference, which often amounts to working out the marginal distributions associated with single vertices in the graph.

(ii) If the graph is known, but the associated parameters are not known, then one can estimate the parameters associated with the conditional distributions encoded by the graph.

(iii) If neither the graph nor the associated parameters are known, then one is often concerned with structural learning, or doing model selection.

In the first two chapters, I develop relevant theoretical results relevant to this last point. I suppose our observed data was generated by a process represented by a directed acyclic graph (DAG) with latent and selection variables. The causal interpretation of such a DAG is described by Spirtes et al. (1993), and Pearl (2000). There may be situations in which data collected from a population represented by a given data-generating process D is such that: i) measurement on some variables are unobserved (latent variables), and ii) some variables have been conditioned on (selection variables). One might think that in this case, though I may not be able to determine the influence of any hidden variables, I could just consider the observed variables and at least correctly represent the independence relations among them. Unfortunately, this is not always the case for DAG models because they are not closed under conditioning or marginalization. This point can be better understood through the following example.

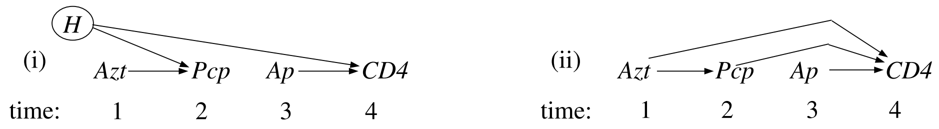


Figure 1.1: (i) A DAG with a latent variable H. (ii) A model search that does not include H may add an extra edge from Azt to $CD4$.

Consider the toy example given in Figure 1.1(i)[1]. Azt is a drug given to AIDS patients to increase their $CD4$ counts. Ap is a drug often given to AIDS patients to treat opportunistic infections. This graph pertains to the hypothetical experiment wherein subjects are randomized to Azt at time 1 and Ap at time 3, and then the outcome, $CD4$ count, is observed some time in the future. Suppose that there are side effects associated with Azt such that some of the patients on Azt develop the opportunistic infection Pcp, but Azt has no effect on $CD4$ count. H refers to a patient's underlying health status, which is not observed. A subject with poor health status may be more likely to develop Pcp (observed at time 2), and she may also be more likely to have a low $CD4$ count. Note that temporal knowledge gives a total ordering on the variables.

The DAG implies the following: $\langle Azt \perp\!\!\!\perp \{Ap, CD4\},\ Ap \perp\!\!\!\perp \{Azt, Pcp\}\rangle$. In particular, note that Azt is marginally independent of $CD4$. Given data generated by this DAG, a search over DAGs containing only the observed variables, and consistent with this time-ordering, would asymptotically find a DAG with an extra edge from Azt to $CD4$ (see Figure 1.1(ii)). From such a search one could draw the incorrect conclusion that Azt influences $CD4$ count. There is no DAG that can represent all of, and only, these independence relations using the observed variables alone. One approach to this problem would be to introduce latent variables into the model. However, introducing latent variables to a model may remove some of the desirable properties of the statistical distributions associated with the graph (see Settimi and Smith (1999) and Geiger et al. (1999)):

[1] The example given in Figure 1.1 is a fictitious experiment based on an observational study analyzed by Hernán et al. (2000).

(a) these models may not be identifiable; the likelihood of the parameters for a specific model may be multi-modal;

(b) inference may be highly sensitive to the assumptions made about the unobserved variables; and

(c) the associated distributions may be difficult to characterize, in particular they may not form a curved exponential family.

If there is detailed background knowledge about the process, then one might use a latent variable model, and exploit this information during the model search process. However, in the absence of background knowledge, we are in a dilemma: including latent variables explicitly can make modelling difficult, particularly when the structure of the graph is not known; *not* including hidden variables can potentially lead to misleading analyses (e.g. extra edges may be introduced to the graph). However, ancestral graphs are a class of graphs that, using only the observed variables, can encode the conditional independence relations given by any data-generating process that can be represented by a DAG with latent and selection variables. More precisely, it is shown in Richardson and Spirtes (2000) that if D is a DAG over the vertex set V with latent variables L and selection variables S, then there exists an ancestral graph $\mathcal{G}$ with vertex set $V \backslash (S \cup L)$ which is Markov equivalent to D on the $V \backslash (S \cup L)$ margin conditional on S. Furthermore, Richardson and Spirtes (2000) have shown that for any ancestral graph $\mathcal{G}$ (DAGs form a subset of ancestral graphs) with latent and selection variables, there are graphical operations corresponding to "marginalization" and "conditioning" such that the resulting graph represents the independence model obtained by taking the set of distributions represented by $\mathcal{G}$ and then integrating out the latent variables and conditioning on the selection variables. The resulting graph is itself an ancestral graph and represents the set of conditional independence relations holding among only the observed variables. Given the selection variables, the associated statistical models retain many of the desirable properties that are associated with DAG models. However, as with DAG models, for any ancestral graph, there are potentially several other graphs that represent the same set of distributions. Such graphs are said to be *Markov equivalent*. Consequently, data cannot distinguish between Markov equivalent graphs.

I define a join operation on ancestral graphs which associates a unique graph with an equivalence class. I also extend the separation criterion (See Definition 2.1.3) for ancestral graphs (which is an extension of d-separation) and prove the pairwise and global Markov properties for joined Markov equivalent maximal ancestral graphs. Andersson et al. (1997) showed that the graph resulting from joining a Markov equivalence class of DAGs is a chain graph. They also characterized the structure of this chain graph and showed that it is Markov equivalent to the original DAGs in the equivalence class. The pairwise Markov property for joining DAGs follows from their finding. Partial characterizations of Markov equivalence classes for ancestral graphs have been obtained using POIPGs and PAGs by Richardson and Spirtes (2002) and Spirtes et al. (1993). A key difference between these authors' works and the present investigation is that the representation given here is guaranteed to include all arrowheads common to every graph in the equivalence class, whereas this is not true in the previous work. In other words, the representation here is guaranteed to be *complete* with respect to arrowheads (see Meek (1995a)). The graphs described here are analogous to the essential graph for DAGs (Andersson et al. (1997)), while previous representations have been analogous to patterns (Verma and Pearl (1991)).

The next two chapters are organized as follows:

Chapter 2 will provide background on ancestral graphs, introduce a simple representation of equivalence classes using joined graphs and prove the pairwise Markov property for joined graphs. Chapter 3 looks at the structure of a special paths, *minimal j-connecting paths* (see Definitions 2.2.2 and 3.1.1) in joined graphs, and presents a proof of the global Markov property for joined graphs.

I now provide some background on DAG models, and characterizing equivalence classes for DAGs.

1.2 Basic Definitions and Concepts

A *graph* $\mathcal{G} = \{V, E\}$ is composed of a set of vertices V and edges E such that $\mathcal{G}$ represents the set of conditional independence holding among the variables in V. A *subgraph* of $\mathcal{G} = \{V, E\}$ is any graph $\mathcal{G}' = \{V', E'\}$ such that $V' \subseteq V$, and every edge in $\mathcal{G}'$ is also in $\mathcal{G}$. The *independence model* associated with a graph is the set of conditional independence relations encoded by the graph.

1.2.1 Vertex Relations

If there is an edge between α and β in the graph $\mathcal{G}$, then α is *adjacent* to (sometimes referred to as *"an adjacency of"*) β and vice versa. Table 1.2.1 provides the terminology for different vertex relations.

Table 1.1: Vertex relation terminology for adjacencies in a graph.

Edge Type	α's relation to β
$\alpha - \beta$	neighbour
$\alpha \leftrightarrow \beta$	spouse
$\alpha \rightarrow \beta$	parent
$\alpha \leftarrow \beta$	child

Note that the three edge types shown in Table 1.2.1 ($-, \leftrightarrow, \rightarrow$) are distinct and should be interpreted as such. I only consider graphs in which there is at most one edge between any two vertices.

1.2.2 Paths

A *path* is a sequence of edges between α and β in $\mathcal{G}$ that contains no repetitions. Consequently, any path can be represented by a sequence of distinct vertices. If there is a directed path from α to β (i.e. $\alpha \rightarrow\rightarrow \ldots \rightarrow \beta$), then α is an *ancestor* of β, and β is a *descendant* of α. Also, this directed path from α to β is called an *ancestral path*. If there is a directed path from α to β with leading undirected edges (i.e. $\alpha - \; - \ldots \rightarrow\rightarrow \beta$), then α is *anterior* to β, and β is a *non-descendant* of α. Note that if α is an ancestor of β, then α is anterior to β as well. If there is a directed path from α to β and $\alpha = \beta$ then the path forms a *cycle*, or the path is termed *cyclic*.

The ancestor set of the vertex set Z is the union of the ancestors of every vertex in Z (except Z itself), and this set is denoted by $an(Z)$. The set of vertices anterior to the set Z is the union of the all vertices anterior to each vertex in Z, (except Z itself), and this set is denoted by $ant(Z)$.

1.3 Directed Acyclic Graphs

Directed acyclic graphs are often used in practice to represent diverse data-generating processes (e.g. acquisition of a disease) because they are relatively easy to work with, and have many desirable properties. DAGs contain only directed edges ($\rightarrow$) with the structural constraint that no path through the graph forms a cycle. In other words, these graphs are acyclic.

1.3.1 Desirable Properties of DAG Models

Scientists often use DAG models because there is a simple causal interpretation of these graphs. Consider the DAG shown in Figure 1.3.1. The $a \leftarrow d$ edge can be interpreted as "d causes a" in the sense that if I were to intervene on variable a then there would be no change in the distribution of d; but, if we were instead to intervene and set d to some constant, then the distribution of a would be modified, and would equal the conditional distribution of a given d (see Pearl (2000)).

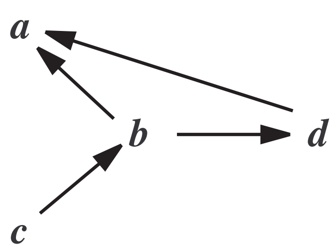


Figure 1.2: An example of a directed acyclic graph (DAG).

As well, there is a natural factorization of the joint density of the vertices in a graph D.

Again, consider the DAG given in Figure 1.3.1. The joint density of D factors in the following way:

$$p(a,b,c,d) = p(c)p(d)p(b|c,d)p(a|b,d)$$

For a general DAG $D = \langle V, E \rangle$, the joint density associated with D factors into a product of the probability of each vertex conditional on its parents:

$$p(V) = \prod_{x \in V} p(x|parents(x))$$

There are also many properties of the statistical models represented by DAGs (with no latent variables) that make inference via graphical models convenient. In particular:

1. These models are identifiable;
2. These models belong to curved-exponential families with well-defined dimension;
3. Methods exist for fitting such models; and
4. Maximum likelihood estimates of the parameters in the statistical model exist.

1.3.2 Independence Model

A graph with vertex set V corresponds to a joint density over the random variables indexed by V. Associated with every graph is a set of marginal and conditional independence relations that follow a "Markov property". The independence model $\mathcal{I}$ over the vertex set V is a set of triples $\langle A, B|C \rangle$, where A, B, C are disjoint subsets of V, with C possibly being the null set. If $\langle A, B|C \rangle \in \mathcal{I}$ then we say that "A is conditionally independent of B given C in $\mathcal{I}$" where A, B, C each denote both a subset of vertices in V and the respective associated set of random variables. I write $A \perp\!\!\!\perp B|C[P]$ to mean "A is independent of B conditional on C in the distribution P" (Dawid (1979)).

A distribution P is said to be *faithful* or *Markov perfect* with respect to an independence model $\mathcal{I}$ if:

$$\langle A, B \mid C \rangle \in \mathcal{I} \quad \Leftrightarrow \quad A \perp\!\!\!\perp B \mid C \; [P].$$

The above condition is known as a "global Markov property". An independence model $\mathcal{I}$ is said to be probabilistic if there is a distribution P that is faithful to $\mathcal{I}$. For the DAG shown in Figure 1.3.1 there are two relations in the associated independence model: $\langle a \perp\!\!\!\perp c | b \rangle$ and $\langle d \perp\!\!\!\perp c | \emptyset \rangle$. See Richardson and Spirtes (2002), Spirtes et al. (1993) and Pearl (1988).

Paths in a graph can informally be thought as a flow of information from one vertex to another. Conditioning on a variable can either "block" a path (stop the flow of information), or "open" the path (allow information to flow). Graphical criteria stipulate when paths are open or blocked. Before stating the d-connection criterion, I provide a few definitions:

A *collider* (also known as a head-to-head node) is a set of three vertices $\{a, b, c\}$ on a path π such that $a \rightarrow b \leftarrow c$. In this example, vertices a and c "collide" at vertex b on the path. This configuration is sometimes referred to as there being "a collider at b on the path π". For DAGs, all other configurations for non-endpoints ($a \rightarrow b \rightarrow c, a \leftarrow b \rightarrow c, a \leftarrow b \leftarrow c$) are termed *non-colliders*. Note that such configurations are relative to a specific path. In the above example, b may be a collider on π, but a non-collider along some other path in the graph. If a and c are not adjacent, then $\{a, b, c\}$ is *unshielded*; conversely, if a and c are adjacent, then $\{a, b, c\}$ is *shielded*.

Definition 1.3.1 *Pearl (1988). A path π between distinct vertices α and β in a DAG D is said to be "d-connecting given Z" (where $\alpha, \beta \notin Z$ and Z may be empty) if the following hold:*

(i) no non-collider on π is in Z, and

(ii) every collider on π is an ancestor of a vertex in Z.

Two vertices α and β are said to be d-separated given Z if there is no path d-connecting α and β given Z.

Open paths path relate to d-connection in that if at least one path between α and β is open given Z, then α and β are d-connected given Z. Blocked paths relate to d-separation in that if every path between α and β is blocked given Z, then α and β are d-separated given Z.

The d-separation property defines the global Markov property for DAGs. Independence models described by DAGs satisfy a pairwise Markov property such that every missing edge corresponds to a conditional independence. In particular, if α and β are not adjacent and

β is not a descendant of α, then α and β are d-separated given the parents of α. (Note that for any given pair of vertices in a DAG, either α is not a descendant of β or β is not a descendant of α, or both). In other words, for every non-adjacent pair of vertices $\langle\alpha, \beta\rangle$ in the graph, there exists some subset of the vertices in the graph Z that d-separates α and β.

1.4 Equivalence classes for DAGs and Chain Graphs

As was mentioned earlier, for a given set of data, there may be many DAGs that are compatible with the data in the sense that the likelihood associated with the statistical model represented by these DAGs are the same. For example, $a \rightarrow b \rightarrow c$, $a \leftarrow b \rightarrow c$ and $a \leftarrow b \leftarrow c$ encode the same conditional independence relation that $a \perp\!\!\!\perp c \mid b$, and hence these three DAGs are indistinguishable via the likelihood. If two graphs encode the same set of conditional independence relations (i.e. the two graphs entail the same independence model), then the graphs are considered to be Markov equivalent.

The following subsections summarize results by Verma and Pearl (1991), Frydenberg (1990) and Andersson et al. (2001), Andersson et al. (1997) and Meek (1995b) respectively.

1.4.1 Patterns (Verma and Pearl (1991))

Verma and Pearl (1991) presented graphical conditions that determine when two DAGs would be Markov equivalent to each other, or would entail the same set of conditional independence relations:

Theorem 1.4.1 *Two DAGs are Markov equivalent if and only if they have the same adjacencies and the same unshielded colliders.*

Theorem 1.4.1 leads to a natural canonical representation of an equivalence class of DAGs: construct a partially directed graph by removing arrowheads from any edge in the DAG that is not participating in an unshielded collider. Verma and Pearl (1991) call the resulting object a "rudimentary pattern" of the causal model. Rudimentary patterns can be identified solely by d-separation because each equivalence class has a unique pattern. Once the rudimentary pattern has been established, there may be additional arrowheads to be added to the pattern which are present in all Markov equivalent DAGs. In this framework, the resulting pattern is called the "completed pattern", and represents the equivalence class for the DAG used to construct the completed pattern in the sense that an edge is directed in the completed pattern if and only if it has the same orientation in all graphs in the equivalence class. Figure 1.3 shows an example of DAGs in an equivalence class, along with the corresponding rudimentary and completed patterns. Note that Meek (1995b) also used patterns to represent equivalence classes of DAGs, but his resulting "completed pattern" differs from that of Verma and Pearl (1991) (see subsection 1.4.4).

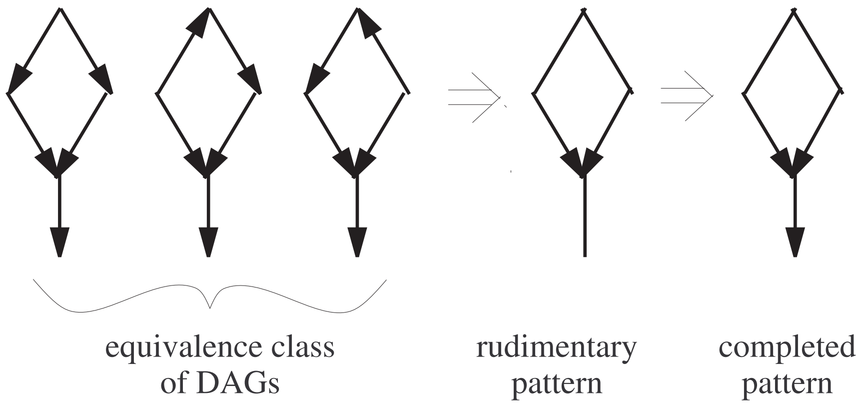


Figure 1.3: An example of a completed pattern, which represents an equivalence class of DAGs.

1.4.2 Chain Graphs (Frydenberg (1990) and Andersson et al. (2001))

Lauritzen and Wermuth (1989) (LWF) introduced a class of graphs called chain graphs for modelling conditional independencies. Frydenberg (1990) and Andersson et al. (2001) independently provided conditions under which two chain graphs are Markov equivalent to each other. To construct a chain graph, the variables relevant to the data-generating process are divided into disjoint sets which are then ordered into a dependence chain that reflects the known causal structure of the process. The variables form the vertex set of the chain graph and edges are drawn according to the following:

i) Two variables α and β are not adjacent if they are conditionally independent given all the other variables that do not come after both of the vertices in the dependence chain; and

ii) Two adjacent variables, α and β, are connected by an undirected edge if they belong to the same set in the dependence chain, and with an arrow from α to β otherwise (where α belongs to a set in the dependence chain that comes before the set to which β belongs).

Thus, chain graphs are composed of directed and undirected edges with the structural constraints outlined above. Each connected subset of the dependence chain is called a *chain component*, and can be identified by removing the directed edges from the graph.

A path through a chain graph that consists of both directed and undirected edges such that all directed edges on the path are pointing in the same direction (e.g. $\alpha - \rightarrow - - \rightarrow - \beta$) is known as a *partially directed* path. Figure 1.4 illustrates examples of *partially directed cycles*, partially directed paths such that the path begins and ends at the same vertex.

Chain graphs do not contain any partially directed cycles consisting of one or more directed edges, a constraint analogous to the acyclic constraint for DAGs. Note that undirected graphs can be viewed as chain graphs in which there is only one chain component; and DAGs can be viewed as chain graphs in which each vertex comprises its own chain component.

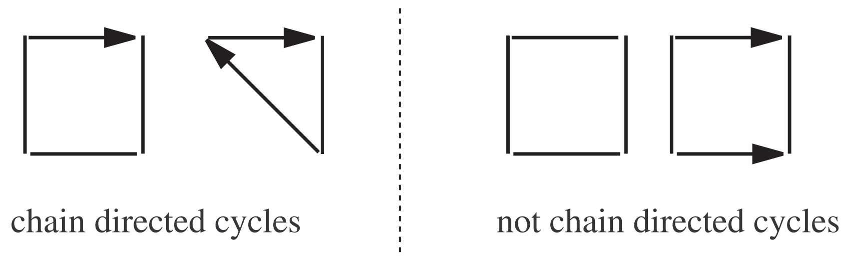


Figure 1.4: Examples of partially directed cycles

For a set of vertices A, in a chain graph $\mathcal{G}$ over vertex set V, the *boundary* of A is: $bd(A) = \{\beta \in V \backslash A \mid \beta \rightarrow \alpha \text{ or } \beta - \alpha \text{ for some } \alpha \in A\}$; and the *closure* of A is:

$cl(A) = bd(A) \cup A$. In other words, the boundary of A corresponds to the union of the parents and neighbours of A, and the closure of A is the union of A with its parents. The vertex set A is called *complete* if for every pair of vertices in A, $\langle \alpha, \beta \rangle$ either $\alpha \rightarrow \beta$, $\beta \rightarrow \alpha$ or $\beta - \alpha$ (i.e. all vertices in A are adjacent to each other). For DAGs, there is a process called "moralization" in which all unshielded colliders are shielded and the arrowheads are removed from the graph. There is an analogous moralization process for chain graphs. The moral graph $\mathcal{G}^m = \langle V, E^m \rangle$ is the underlying undirected graph (i.e. all arrowheads are removed) where the boundary, with respect to $\mathcal{G}$, of every chain component is made complete. Figure 1.5 shows an example of a chain graph $\mathcal{G}$, identifies the chain components of $\mathcal{G}$, and shows the corresponding moral graph $\mathcal{G}^m$.

Before stating the main result of Frydenberg (1990), I make one more note: if P is the measure over V for an undirected graph, and if P has a positive density, then the following conditional independence axiom holds (from Section 3. of Andersson et al. (2001), also axiom (CI4) in Frydenberg (1990)):

(CI5) $A \perp\!\!\!\perp B \mid D \cup C \, [P]$ and $A \perp\!\!\!\perp C \mid D \cup B \, [P]$ implies $A \perp\!\!\!\perp B \cup C \mid D \, [P]$

An induced subgraph $\mathcal{G}_A$ of a graph $\mathcal{G}$ with vertex set V is that graph comprised of the vertices in A and containing all of the edges in $\mathcal{G}$ involving all and only the vertices in A. Let $\phi(\alpha)$ consist of all vertices that can be reached from α by a partially directed path. The probability measure P is then said to be:

(P) *Pairwise G-Markovian* if $\alpha \perp\!\!\!\perp \beta \mid [\, V \setminus \phi(\alpha)] \setminus \{\alpha, \beta\} \; [P]$ whenever $\beta \notin \phi(\alpha)$ and β and α are not adjacent

(L) *Local G-Markovian* if $\alpha \perp\!\!\!\perp [V \setminus \phi(\alpha)] \setminus cl(\alpha) \mid bd(\alpha) \; [P]$ for all α

(G) *Global G-Markovian* if $A \perp\!\!\!\perp B \mid C \; [P]$ whenever C separates A and B in $(\mathcal{G}_{an(A \cup B \cup C)})^m$

where $an(A \cup B \cup C)$ is the smallest anterior set containing $A \cup B \cup C$ (see Section 1.2.2). Frydenberg (1990) proved that if (CI5) holds for any P and any chain graph $\mathcal{G}$, then P is pairwise, local and global G-Markovian. He also showed that associated with chain graphs is a natural factorization of the associated joint density.

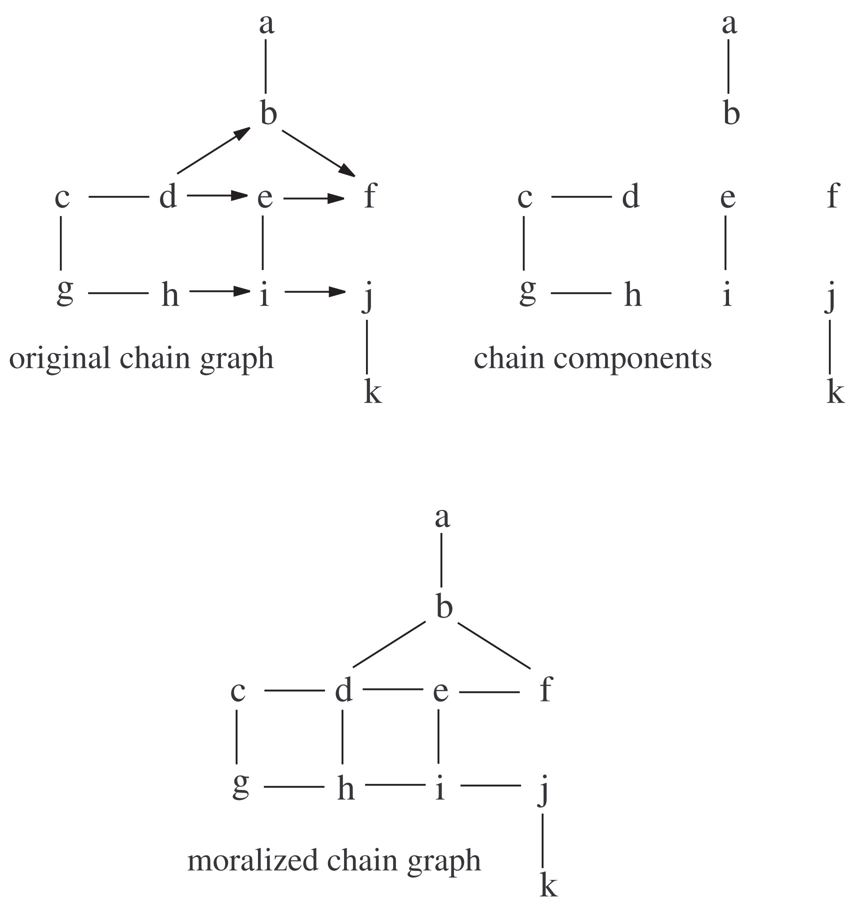


Figure 1.5: Example of Chain Graph

A triple (α, B, β) is called a *complex* in $\mathcal{G}$ if B is a connected subset of a chain component τ and α and β are two non-adjacent vertices in $bd(\tau) \cap bd(B)$. Further, a complex (α, B, β) is a *minimal* complex if $B = B'$ whenever B' is a subset of B, and (α, B', β) forms a complex. It can easily be verified that the general structure of a minimal complex is like that shown in Figure 1.6.

Definition 1.4.1 *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two chain graphs with the same adjacencies such that all undirected edges in $\mathcal{G}_2$ are undirected in $\mathcal{G}_1$, and there is at least one edge that is directed in $\mathcal{G}_2$ but is undirected in $\mathcal{G}_1$. Then $\mathcal{G}_1$ is larger than $\mathcal{G}_2$.*

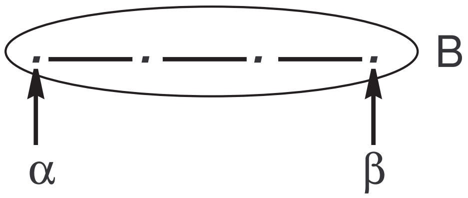


Figure 1.6: Structure of a Complex in a Chain Graph

Thus, if $\mathcal{G}$ and $\mathcal{G}'$ are two chain graphs with the same adjacencies, and $\mathcal{G}'$ is larger than $\mathcal{G}$, then $\mathcal{G}$ and $\mathcal{G}'$ have the same minimal complexes if and only if any minimal complex in $\mathcal{G}$ is a complex in $\mathcal{G}'$. Given this finding, Frydenberg (1990) (and Andersson et al. (1997) independently) proved that:

Theorem 1.4.2 *(Frydenberg (1990) and Andersson et al. (1997)) Two chain graphs have the same Markov properties if and only if they have the same underlying undirected graph and the same minimal complexes.*

Andersson et al. (2001) (AMP) introduced an alternative interpretation of chain graphs, and presented a corresponding Markov property for these graphs. Before presenting their Markov property, I make the following definitions:

Definition 1.4.2 *A triplex in the chain graph $\mathcal{G}$ is an ordered pair $\langle\{a,b\},\ c\rangle$ such that $\{a,b,c\}$ occur in one of the configurations shown in Figures 1.7(a)-(c) as an induced subgraph of $\mathcal{G}$.*

Another configuration that occurs in chain graphs is shown in Figure 1.7*(d)*. Such configurations are termed *2-biflags*. I now define augmentation for AMP chain graphs. Examples of augmented chain graphs are provided in Andersson et al. (2001).

Definition 1.4.3 *The augmented graph $\mathcal{G}^a$, derived from chain graph $\mathcal{G}$ is defined to be the undirected graph obtained by augmenting all triplexes and 2-biflags in $\mathcal{G}$, as shown in Figures 1.8(a) and 1.8(b) respectively, then converting all remaining directed edges of $\mathcal{G}$ into undirected edges.*

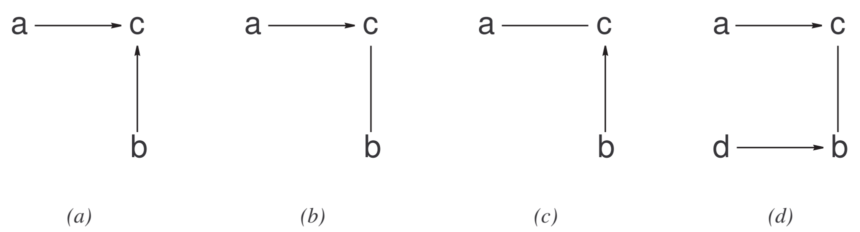


Figure 1.7: *(a), (b), (c)* Examples of Triplexes in Chain Graphs, *(d)* Example of a 2-bi-flag

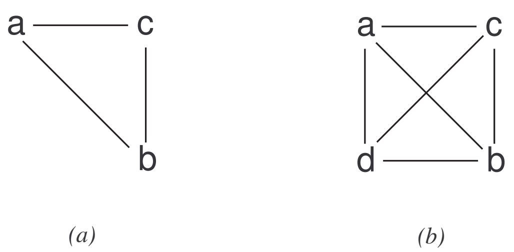


Figure 1.8: Augmentation for Triplexes and 2-biflags

The AMP global Markov property for chain graphs can be stated as follows:

Definition 1.4.4 *(Andersson et al. (2001)) Let $\mathcal{G}$ be a chain graph. A probability measure P on $\mathbf{X}$ is said to be AMP global G-Markovian if $A \perp\!\!\!\perp B \mid S\ [P]$ whenever S separates A and B in $\mathcal{G}[A\dot{\cup}B\dot{\cup}S]^a$.*

Under the AMP formulation of chain graphs, Andersson et al. (2001) show that:

Theorem 1.4.3 *(Andersson et al. (2001)) Two chain graphs have the same Markov properties if and only if they have the same underlying undirected graph and the same triplexes.*

Because DAGs are a special instance of chain graphs, these results also apply to DAGs and state when two DAGs are Markov equivalent to each other. Furthermore, they show that the AMP formulation of chain graphs coincide with the LWF formulation for undirected graphs and for DAGs.

1.4.3 The Essential Graph (Andersson et al. (1997))

Like Verma and Pearl (1991), Andersson et al. (1997) (AMP) also examined equivalence classes for DAG models and presented a characterization of the equivalence class. They showed that each Markov equivalence class is uniquely determined by a single chain graph, called the "essential graph", and that this graph is a chain graph that is Markov equivalent to all the DAGs in the original equivalence class. This chain graph is not the same completed pattern constructed by Verma and Pearl (1991). Andersson et al. (1997) also presented a proof of Theorem 1.4.1 that is different from that of Verma and Pearl (1991).

Definition 1.4.5 *The "essential graph" $\mathcal{D}$ with DAG D is the graph*

$$\mathcal{D} \;:=\; \cup\,(D' \mid D' \sim D)$$

i.e. the smallest graph larger than every $D' \in [D]$.

In other words, an arrowhead is present at a vertex in $\mathcal{D}$ if and only if that arrowhead is present at that vertex in every member of the equivalence class. Figure 1.9 shows an example of an equivalence class of DAGs, along with the corresponding essential graph.

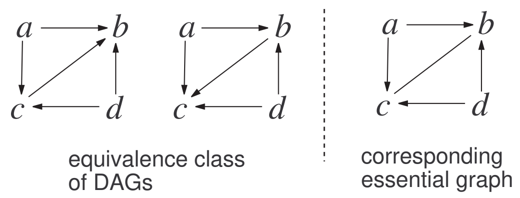


Figure 1.9: An example of an essential graph, which represents an equivalence class of DAGs.

An arrow $\alpha \to \beta$ is said to be *strongly protected* in $\mathcal{G}$ if the edge occurs in $\mathcal{G}$ in at least one of the following four configurations shown in Figure 1.10 as an induced subgraph of $\mathcal{G}$.

Using these concepts, Andersson et al. (1997) provided a full characterization of the essential graph:

Theorem 1.4.4 *A graph $\mathcal{G} \equiv \langle V, E\rangle$ is equal to the essential graph $\mathcal{D}$ for some DAG D if and only if the following four conditions hold:*

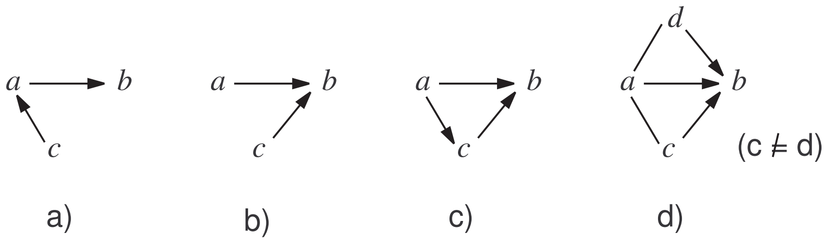


Figure 1.10: The edge $\alpha \to \beta$ is strongly protected in a DAG if it appears in at least one of these configurations as an induced subgraph.

(i) $\mathcal{G}$ *is a chain graph;*

(ii) for every chain component τ *of* $\mathcal{G}$, $\mathcal{G}_\tau$ *is chordal;*

(iii) the configuration $a \to b -$ *does not occur as an induced subgraph of* $\mathcal{G}$*; and*

(iv) every arrow $a \to b$ *is strongly protected in* $\mathcal{G}$.

1.4.4 Orientation Rules (Meek (1995b))

Meek (1995b), Andersson et al. (1997) and Chickering (1995) independently addressed the question of how to construct the essential graph for a DAG equivalence class when one member (graph) of the class is known by applying a number of orientation rules to the known member of the equivalence class. In this section, we discuss the set of orientation rules introduced by Meek (1995b), who had two questions of interest:

(i) Given a data sample, is there a causal explanation (i.e. a DAG) that is consistent with a set of background knowledge that is supported by the data?

(ii) Given that a causal explanation exists, what are the causal relationships that are common to every causal explanation that is supported by the data?

Here I present the algorithm for answering question (ii). A *complete* causal explanation of an independence model $\mathcal{I}$ is a DAG D if and only if the conditional independence relations encoded by D are exactly those in $\mathcal{I}$. Then, there are two phases to this algorithm:

I. Examine the independence statements in the independence model and try to construct the pattern of some DAG D. Let Π_I be the result of this phase.

II. Find a partially directed graph whose adjacencies are the same as any complete causal explanation for $\mathcal{I}$ and whose edges are directed if and only if every complete causal explanation for $\mathcal{I}$ (i.e. every member of the equivalence class for D) has the edge oriented as such.

Phase I consists of the following steps:

1. Let $\Pi_I = (V, \emptyset)$, (i.e. the vertex set of D with no edges between any pair of vertices).
2. Add an undirected edge between A and B in Π_I if and only if A and B cannot be d-separated in D.
3. Add arrowheads at B in Π_I such that $A \rightarrow B$ and $B \leftarrow C$ if and only if $\langle A, B, C\rangle$ forms an unshielded collider (i.e. A and C are d-separated by some set not containing B).

Phase II consists of successively applying a number of orientation rules to pattern Π_I where the orientation rules are as shown in Figure 1.11. The orientation rules given by Meek (1995b) provide a method of constructing the essential graph without having to first identify every member of the equivalence class.

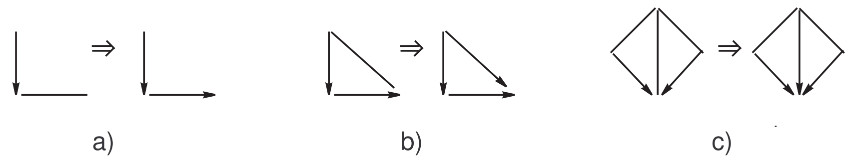


Figure 1.11: Orientation rules for constructing essential graph

Recall from Section 1.1 that DAGs are not rich enough to encode the conditional independence relations holding among the observed variables of a data-generating process that can be represented by a DAG with latent and selection variables. In Chapter 2 I provide background for a broader class of graphs, ancestral graphs, which can encode such models. It will be shown that Markov equivalence of ancestral graphs require all graphs in the equivalence class have the same adjacencies and same unshielded colliders. However, there is an additional requirement involving certain shielded vertices in the graphs. I also introduce a simple representation for equivalence classes of ancestral graphs, that is analogous to the essential graph for DAGs.

Chapter 2

JOINED GRAPHS

In this chapter I review ancestral graphs (Section 2.1), and define a new class of graphs: joined graphs (Section2.2). In Section 2.3 I introduce a simple representation of an equivalence class of Markov equivalent maximal ancestral graphs, and in the remaining sections of the chapter I work towards characterizing the structure of joined graphs. The first main result of this thesis, the Pairwise Markov property for joined graphs (Theorem 2.6.1), is in this chapter as well. An outline of this proof is given in Section 2.6 (see Figure 2.22).

2.1 Background: Ancestral graphs (Richardson and Spirtes (2000))

The basic motivation for developing ancestral graphs is to enable one to focus on the independence structure over the observed variables that results from the presence of latent variables without explicitly including latent variables in the model. Permitting bi-directed ($\leftrightarrow$) edges in the graph allows one to graphically represent the existence of an unobserved common cause of observed variables. For Figure 1.1(i) this corresponds to removing H from the graph and adding a bi-directed edge between Pcp and $CD4$. Undirected edges (—) are also introduced to represent unobserved selection variables that have been conditioned on rather than marginalized over. However, interpreting ancestral graphs is not so straightforward. Richardson and Spirtes (2002) provide a detailed discussion on the interpretation of edges in an ancestral graph. Further details of the basic definitions and concepts presented here can also be found in Richardson and Spirtes (2000).

Definition 2.1.1 *A graph, which may contain undirected (—), directed ($\rightarrow$) and bi-directed edges ($\leftrightarrow$) is ancestral if:*

(a) there are no directed cycles;

(b) *whenever an edge $x \leftrightarrow y$ is in the graph, then x is not an ancestor of y, (and vice versa);*

(c) *if there is an undirected edge $x — y$ then x and y have no spouses or parents.*

Conditions *(a)* and *(b)* may be summarized by saying that if x and y are joined by an edge and there is an arrowhead at x, then x is *not* an ancestor of y; this is the motivation for the term 'ancestral'. Note that by *(c)*, the configurations $\rightarrow \gamma —$ and $\leftrightarrow \gamma —$ never occur in an ancestral graph. For ancestral graphs, a non-endpoint vertex v on a path is said to be a *collider* if two arrowheads meet at v, i.e. $\rightarrow v \leftarrow$, $\leftrightarrow v \leftrightarrow$, $\leftrightarrow v \leftarrow$ or $\rightarrow v \leftrightarrow$; all other non-endpoint vertices on a path are *non-colliders*, i.e. $— v —$, $— v \rightarrow, \rightarrow v \rightarrow, \leftarrow v \rightarrow, \leftrightarrow v \rightarrow$. These definitions of collider and non-collider are direct extensions of the corresponding definitions for DAGs.

2.1.1 Marginalization & Conditioning

This section discusses marginalization and conditioning for an independence model. For processes that can be represented by DAGs with latent and selection variables, there are graphical transformations that correspond to marginalizing and conditioning such that the resulting graph contains only the observed variables of the process, and the associated independence model contains all and only the conditional independence relations holding among the observed variables.

Let us consider the example illustrated in Figure 2.1(a). Let S represent smoking status, Yf represent yellowed fingers, and LC represent lung cancer status. In this example, a person who smokes may be more likely to have yellowed fingers due to nicotine stains, and she may also be more likely to develop lung cancer. Suppose I have data generated from this DAG and I are interested in the relationship between having yellowed fingers and developing lung cancer. In the absence of information about smoking status, tests of association between Yf and LC would be statistically significant, given a sufficient sample size. Graphically, marginalizing over smoking status corresponds to removing all edges out of smoking status, adding a bi-directed edge between Yf and LC, and removing S from the graph. According to the original DAG, conditional on smoking status, Yf and LC

are independent. Graphically, conditioning on S corresponds to breaking the edges out of smoking status and removing S from the graph. The graphs resulting from marginalizing over S and conditioning on S are shown in Figures 2.1(b) and 2.1(c) respectively. However, the graphical transformations for marginalizing and conditioning can be more complicated.

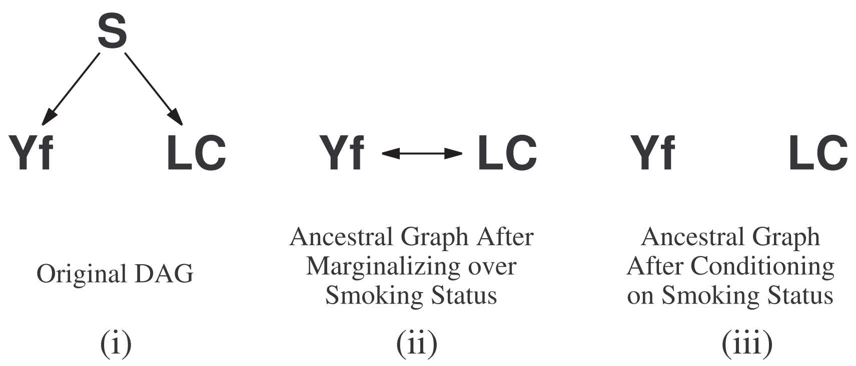


Figure 2.1: An Example of Marginalizing and Conditioning on Variables in a DAG Model. See text for further explanation.

Definition 2.1.2 *Graph $\mathcal{G}[^{S}_{L}$ has vertex set $V \setminus (S \dot{\cup} L)$, and edges specified as follows:*

If α, β, are such that $\forall Z$, with $Z \subseteq V \setminus (S \cup L \cup \{\alpha, \beta\})$,

$$\langle \{\alpha\}, \{\beta\} | Z \cup S \rangle \notin \mathfrak{I}_m(\mathcal{G}),$$

$$\textit{and} \left\{ \begin{array}{c} \alpha \in \mathrm{ant}_{\mathcal{G}}(\{\beta\} \cup S);\ \beta \in \mathrm{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \\ \alpha \notin \mathrm{ant}_{\mathcal{G}}(\{\beta\} \cup S);\ \beta \in \mathrm{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \\ \alpha \in \mathrm{ant}_{\mathcal{G}}(\{\beta\} \cup S);\ \beta \notin S) \\ \alpha \notin \mathrm{ant}_{\mathcal{G}}(\{\beta\} \cup S);\ \beta \notin \mathrm{ant}_{\mathcal{G}}(\{\alpha\} \cup S) \end{array} \right\} \textit{then} \left\{ \begin{array}{c} \alpha - \beta \\ \alpha \leftarrow \beta \\ \alpha \rightarrow \beta \\ \alpha \leftrightarrow \beta \end{array} \right\} \textit{in } \mathcal{G}[^{S}_{L}.$$

Consider the DAG shown in Figure 2.2(a) with latent variables l_1 and l_2, and selection variable s. If I were to first marginalize over l_1 and l_2, then directed edges from a to b and c respectively would be added to the graph, as well as an edge from b to c (see Figure 2.2(b)). In the original DAG, there is anterior path from a to b ($\langle a, l_1, b \rangle$) that passes through l_1 so the $a \to b$ edge is added. Additionally, in the original DAG, there is an anterior path from a to c ($\langle a, l_1, b, s, d, c \rangle$) that passes through l_1, and $b \leftarrow l_2 \to c$ occurs in $\mathcal{G}$ so the edge $a \to c$ is added to the graph. The edge $b \to c$ is added to $\mathcal{G}$ because the configuration $b \leftarrow l_2 \to c$ occurs in $\mathcal{G}$ and there is an anterior path from b to c in $\mathcal{G}$: $\langle b, s, d, c \rangle$. Subsequent conditioning on s corresponds to converting the edge between a and b to an undirected edge, and removing s from the graph along with all edges into and out of s as shown in Figure 2.2(d). If we were to condition on s first, the graph anterior to s is moralized, all edges anterior to s are converted to undirected edges, and s is removed from the graph along with edges directly into and out of s as shown in Figure 2.2(c). Subsequent marginalization over l_1 and l_2 corresponds to adding a directed edge from a to c because there is an anterior path from a to c ($\langle a, l_1, l_2, c \rangle$) in the graph obtained after conditioning on s. The diagram in Figure 2.2 commutes in the sense that the graph resulting from first marginalizing over $\{l_1, l_2\}$ and then conditioning on s is the same graph resulting from first conditioning on s and then marginalizing over $\{l_1, l_2\}$. Furthermore $\mathcal{G}[^S_L$, the graph resulting from marginalizing over latent variables and conditioning on selection variables is an ancestral graph, so ancestral graphs are closed under conditioning and marginalization.

2.1.2 M-connection

Richardson and Spirtes (2000) applied a natural extension of Pearl's d-separation criterion to ancestral graphs:

Definition 2.1.3 *In an ancestral graph, a path π between α and β is said to be m-connecting given Z if the following hold:*

(i) No non-collider on π is in Z; and
(ii) Every collider on π is an ancestor of a vertex in Z.

Two vertices α and β are said to be m-separated given Z if there is no path m-connecting α and β given Z.

Definition 2.1.3 is an extension of the original definition of d-separation for DAGs in that

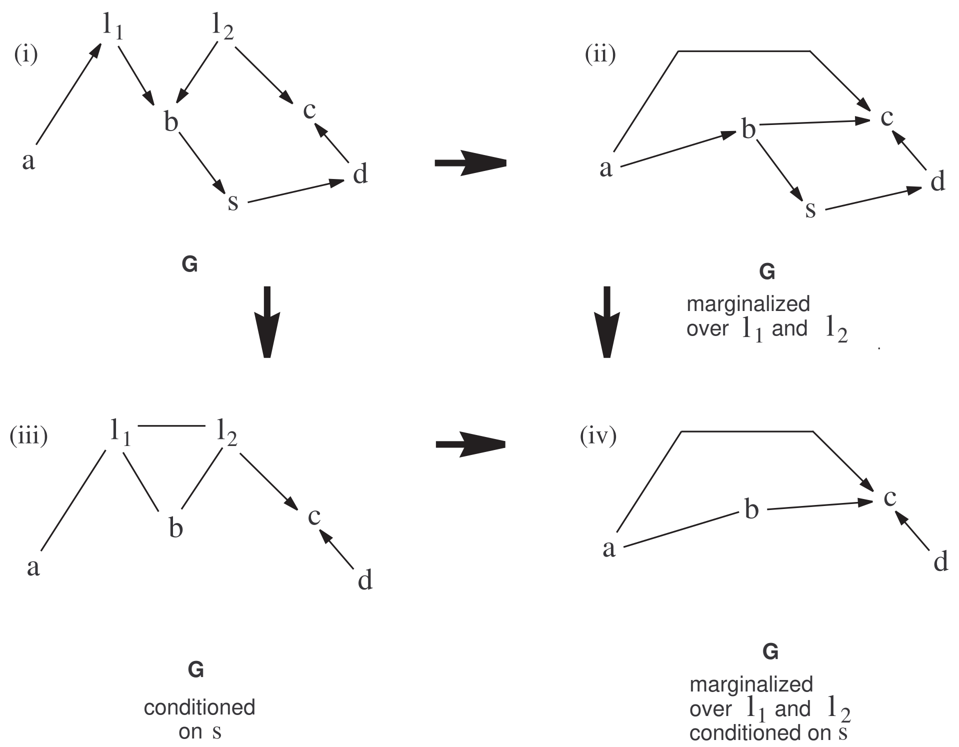


Figure 2.2: A more complicated example of marginalizing and conditioning. See text for further explanation.

the notions of 'collider' and 'non-collider' now allow for bi-directed and undirected edges. m-separation defines the global Markov property for ancestral graphs:

Definition 2.1.4 *(Richardson and Spirtes (2000)) If for all disjoint sets X, Y, Z (Z may be empty), X is m-separated from Y given Z in the ancestral graph $\mathcal{G}$, then $X \perp\!\!\!\perp Y \mid Z\ [P]$, then P is said to satisfy the global Markov property with respect to $\mathcal{G}$.*

If two ancestral graphs have the same vertex sets and m-separation relations, then they are Markov equivalent.

Definition 2.1.5 *Two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be Markov equivalent if for all disjoint sets A, B, Z (where Z may be empty), A and B are m-separated given Z in $\mathcal{G}_1$ if and only if A and B are m-separated given Z in $\mathcal{G}_2$.*

Independence models described by DAGs satisfy pairwise Markov properties such that every missing edge corresponds to a conditional independence relation. In general, this property does not apply to ancestral graphs. For example, there is no set which m-separates γ and δ in the graph in Figure 2.3(a), which motivates the following definition:

Definition 2.1.6 *An ancestral graph $\mathcal{G}$ is said to be "maximal" if, for every pair of non-adjacent vertices α, β there exists a set $Z(\alpha, \beta \notin Z)$, such that α and β are m-separated conditional on Z.*

These graphs are termed *maximal* in the sense that no additional edge may be added to the graph without changing the associated independence model. It has been shown in Richardson and Spirtes (2000) that if an ancestral graph is not maximal, then there exists at least one pair of non-adjacent vertices $\{\alpha, \beta\}$, for which there is an "inducing path" between α and β where:

Definition 2.1.7 *An inducing path π is a path in an ancestral graph such that each non-endpoint vertex is a collider, and an ancestor of at least one of the endpoints.*

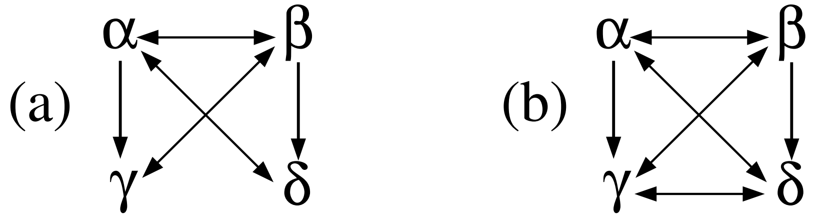


Figure 2.3: (a) The path $\langle\gamma, \beta, \alpha, \delta\rangle$ is an example of an inducing path in an ancestral graph. (b) A maximal ancestral graph Markov equivalent to (a).

Figure 2.3(a) shows an example of a non-maximal ancestral graph. By adding a bi-directed edge between γ and δ, the graph can be made maximal, as shown in Figure 2.3(b). Richardson and Spirtes (2000) show that this is true in general.

One of the key differences between DAGs and ancestral graphs is that there are some shielded colliders in ancestral graphs $\mathcal{G}$ that must be present in any other ancestral graph Markov equivalent to $\mathcal{G}$. In DAGs, there are also shielded colliders that are common to all members of the equivalence class, but the difference between Markov equivalence for DAGs and ancestral graphs is that having the same adjacencies and unshielded colliders is sufficient for Markov equivalence in DAGs. See Figure 2.4 for an example of a set of Markov equivalent DAGs with a shielded collider in all members of the class.

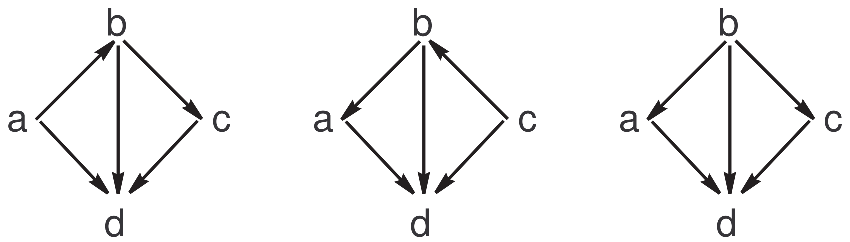


Figure 2.4: A set of Markov equivalent DAGs in which $\langle a, d, b\rangle$ and $\langle c, d, b\rangle$ are shielded colliders that are present in all DAGs in the corresponding equivalence class.

Discriminating paths are useful for identifying which shielded colliders (and non-colliders) are required for ancestral graphs to be Markov equivalent:

Definition 2.1.8 *$U = \langle x, q_1, q_2, \ldots, q_p, \beta, y\rangle$ is a discriminating path for β in an ancestral graph $\mathcal{G}$ if and only if:*

(i) U is a path between x and y with at least three edges,

(ii) U contains $\beta, \beta \neq x, \beta \neq y$,

(iii) β is adjacent to y on U, x is not adjacent to y, and

(iv) For every vertex $q_i, 1 \leq i \leq p$ on U, excluding x, y, and β, q_i is a collider on U and q_i is a parent of y.

Given a set Z, if Z does not contain all $q_i, 1 \leq i \leq p$, then the path $\langle x, q_1, \ldots, q_j, y\rangle$ is m-connecting where $q_j \notin Z$ and $q_i \in Z$ for all $i < j$. If Z separates x and y then Z

contains $\{q_1, \ldots, q_p\}$. Thus, if β is a collider on the path U in the graph $\mathcal{G}$, then $\beta \notin Z$ if Z m-separates x and y. Conversely, if β is a non-collider on the path U then β is a member of any set that m-separates x and y, and β is a non-collider on U in any graph Markov equivalent to $\mathcal{G}$ containing U. Consequently, in any graph $\mathcal{G}^*$ Markov equivalent to $\mathcal{G}$ that contains the discriminating path U, β is a collider on U in $\mathcal{G}^*$ if and only if β is a collider on U in $\mathcal{G}$. In other words, β is "discriminated" to be either a collider or a non-collider on the path U in any graph Markov equivalent to $\mathcal{G}$ in which U forms a discriminating path, even though it is shielded on the path and in the graph. The paths $\langle x, q, \beta, y\rangle$ in $\mathcal{G}_1$ and $\mathcal{G}_2$ from Figure 2.8 are examples of discriminating paths for β.

Lemma 2.1.1 *Let $U = \langle x, q_1, \ldots, q_p, \beta, y\rangle$ form a discriminating path for β in some ancestral graph $\mathcal{G}$. If β is a non-collider on U, then $\beta \to y$ in $\mathcal{G}$.*

Proof:

If β is a non-collider on U, and $\beta ? \to q_p \to y ? - ? \beta$ occurs on U then $\beta \to y$ in $\mathcal{G}$ since otherwise the graph would not be ancestral. $\square$

Considering shielded colliders is not important in determining Markov equivalence for DAGs because (as can easily be verified) such paths always discriminate non-colliders (see Figure 2.8).

Definition 2.1.9 *A "collider path" in an ancestral graph $\mathcal{G}$ is a path such that every vertex, except the endpoints, is a collider on that path.*

From the definition of a discriminating path, the sub-path of U from x to β forms a *collider path.* So referring to $\mathcal{G}_1$ and $\mathcal{G}_2$ in Figure 2.8, the path $\langle x, q, \beta\rangle$ is a collider path (and in fact, in these examples, $\langle x, q, \beta, y\rangle$ forms a collider path too).

2.1.3 Relating Ancestral Graphs to Chain Graphs

DAGs can be viewed as chain graphs in which each chain component consists of a single vertex. DAGs form a subset of ancestral graphs, but only some ancestral graphs can be viewed as chain graphs. Figure 2.5 shows which ancestral graphs can be interpreted via LWF chain graphs and/or AMP chain graphs. Note that if a chain graph is ancestral, then it is Markov equivalent under AMP and LWF since such a graph contains no arrowheads meeting undirected edges. The subset of graphs for which AMP chain graphs, LWF chain graphs and ancestral graphs are Markov equivalent are recursive "causal" models.

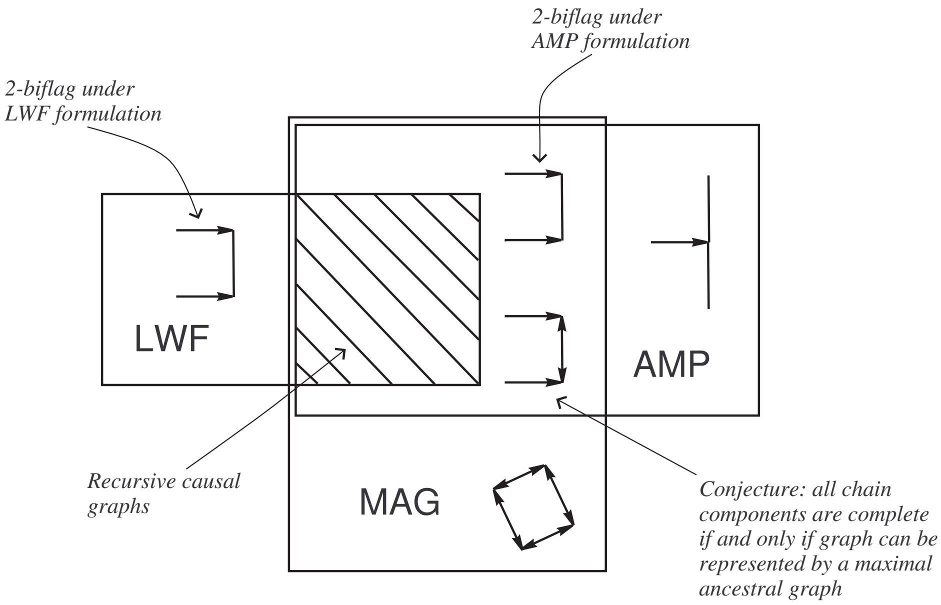


Figure 2.5: Relating Ancestral Graphs to Chain Graphs. The sets here represent independence models. 2-biflags are not ancestral, but there exists ancestral graphs that encode the conditional independence relation represented by such a structure under the AMP chain graph formulation. No such interpretation can be made for 2-biflags in LWF chain graphs.

Table 2.1 summarizes and compares the theoretical properties of many of the graphs reviewed in this thesis (and of undirected graphs).

2.1.4 Characterization of Markov Equivalence

As with DAG models, there is often more than one ancestral graph that can represent a given set of conditional independence statements. Let $[\mathcal{G}]$ represent a Markov equivalence class of ancestral graphs. An interesting question is:

Question 2.1.1 *Which edges or structures are common to all ancestral graphs in* $[\mathcal{G}]$ *?*

Table 2.1: Summary table of theoretical properties of different classes of graphs.

	Undirected graphs	**DAGs**	**LWF Chain graphs**	**AMP Chain graphs**	**Maximal Ancestral graphs**
Markov Property (MP)	global, local, pairwise	global, local, pairwise, well-numbered	global $\Leftrightarrow$ LWF block-recursive MP	global $\Leftrightarrow$ AMP block-recursive MP	global $\Leftrightarrow$ local, pairwise
Density Factorization	factorization according to cliques	factorization recursive, $\prod f(\nu\|pa(\nu))$	$\prod f(\tau\|pa(\tau))$, $f(\tau\|pa(\tau))$ is product of potentials	$\prod f(\tau\|pa(\tau))$, $(\beta_\tau)_{ij} = 0$, *iff* (Gaussian) no $i \to j, \vee\{i,j\}$	none, in general
Markov Equivalence	*iff* $\mathcal{G}_1 = \mathcal{G}_2$	*iff* $\mathcal{G}_1, \mathcal{G}_2$ have same adjacencies & unshielded colliders	*iff* $\mathcal{G}_1, \mathcal{G}_2$ have same adjacencies & complexes	*iff* $\mathcal{G}_1, \mathcal{G}_2$ have same adjacencies & triplexes	*iff* $\mathcal{G}_1, \mathcal{G}_2$ have same adjacencies & discriminating paths with order
Equivalence Class	$\mathcal{G}^* = \mathcal{G}$	$\mathcal{D}$	$\mathcal{G}_\infty$	$\mathcal{G}^*$	$sup[\mathcal{G}]$
Pathwise Separation	separation	d-separation	c-separation (Studeny)	p-separation	m-separation
Completeness	of global MP	of global MP (via d-separation)	of global MP (Studeny et. al)	of global MP (via p-separation)	of global MP (from DAG completeness)
Statistical Inference	no explicit MLE's in Gaussian & multinomial cases	easy due to recursive factorization	partial results (e.g. "BIFROST")	Gaussian case: combine "seemingly unrelated regressions" & "covariance selection"	Gaussian case (numerical methods)
Closed under Marginalization, Conditioning	yes	no	no	no	yes
Interpretation	equilibrium distribution of temporal feedback processes	fully observed causal process	sequence of equilibrium distributions of fully observed feedback		partially observed causal process

This question is of central importance because:

(a) It tells us how much we can learn about causal structure from conditional independence relations over the observed variables.

(b) It would allow us to construct a 'canonical' representative from the Markov equivalence class $[\mathcal{G}]$ which would facilitate searching for a good model.

Spirtes and Richardson (1997) proved the following result for ancestral graphs:

Theorem 2.1.1 *(Markov Equivalence) Two maximal ancestral graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if:*

(i) $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies;
(ii) $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders; and
(iii) If $\mathcal{U}$ forms a discriminating path for β in $\mathcal{G}_1$ and $\mathcal{G}_2$, then β is a collider on U in $\mathcal{G}_1$ if and only if it is a collider on U in $\mathcal{G}_2$.

Theorem 2.1.1 is strengthened in Appendix A by weakening condition (iii). Verma and Pearl (1991) tried to characterize what they called "embedded causal models" in which only a subset of the variables in some process are observed. Hybrid graphs, graphs in which variables that are non-causally but directly correlated are connected by a bi-directed edge, were used to represent embedded causal models. Verma and Pearl (1991) applied natural extensions of constructions for *rudimentary* and *completed patterns* for DAGs to represent equivalence classes of embedded causal models (see Section 1.4). Verma and Pearl (1991) then stated in Theorem 2 of their paper that: *Two embedded causal models are equivalent if and only if they have the same pattern.*

However, there is a flaw in their construction: as we have seen, certain shielded colliders must be retained when determining Markov equivalence of hybrid graphs. Consider the hybrid graphs depicted in Figure 2.6. The graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies and unshielded colliders; hence they entail the same completed pattern (Figure 2.6(c)). But these graphs are not Markov equivalent: for DAGs, if $\langle x, y, z \rangle$ forms a collider, then by the absence of bi-directed edges in the DAG, one can infer that x and z are parents of y; hence, $\{x, z\}$ are non-colliders on the path, but this is not the case for hybrid graphs (or for ancestral graphs).

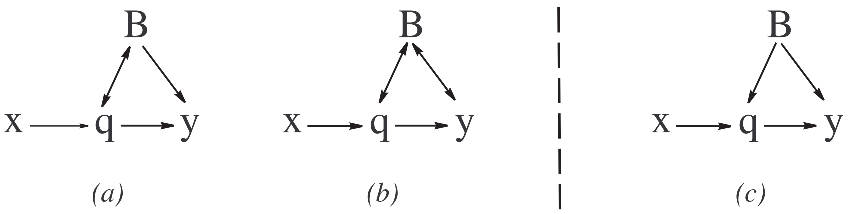


Figure 2.6: *(a),(b)* Two hybrid graphs with same adjacencies and unshielded colliders, but are not Markov equivalent. *(c)* completed pattern for graphs shown in *(a)* and *(b)* are the same.

2.1.5 Partially Ancestral Graphs (Richardson and Spirtes (2002))

Richardson and Spirtes (2002) used graphs that are slightly more complicated than those presented in this thesis to represent equivalence classes of ancestral graphs: *partial ancestral graphs* (PAGs). The key difference between PAGs and joined graphs (see Section 2.2) is that PAGs make the distinction between edges that are present in all ancestral graphs that gave rise to a PAG and those edges in which there is at least one ancestral graph in the equivalence class has an arrowhead at a vertex on a given path, and at least one other ancestral graph in the same equivalence class has a tail at that same vertex on the analogous path.

Given a maximal ancestral graph $\mathcal{G}$, the Fast Causal Inference (FCI) algorithm, Spirtes et al. (1993, 1995, 1999), constructs a PAG, $\mathbf{\Phi}(\mathcal{G})$ to represent a set of graphical features, such as the presence or absence of arrowheads at the ends of particular edges. It is proved by Spirtes et al. (1999) that

$$\mathcal{G}' \in [\mathcal{G}] \quad \text{if and only if} \quad \mathcal{G}' \text{ has all the features given by } \mathbf{\Phi}(\mathcal{G}).$$

The FCI algorithm is usually stated as taking as input an 'oracle' for deciding d-separation relations in a DAG with latent variables (L) and selection variables (S). (Where an 'oracle' is just a black box which gives you answers to questions, in this case whether d-separation relations hold.) However, this is essentially equivalent to taking as input a maximal ancestral graph, or an oracle for deciding m-separation relations in a maximal ancestral graph because any DAG $\mathcal{G}$ with latent variables (L) and selection variables (S) can be represented by the maximal ancestral graph $\mathcal{G}[^S_L$ and vice versa. Furthermore, the m-separation relations in $\mathcal{G}[^S_L$ are equivalent to the d-separation relations that hold in the DAG $\mathcal{G}$ among the observed variables. However, the FCI algorithm is not guaranteed to give a full answer to (a) since it is not known whether the set of features represented by $\mathbf{\Phi}(\mathcal{G})$ is *complete* in the following sense:

> *If some feature is not included in the set represented by $\mathbf{\Phi}(\mathcal{G})$ then there exist graphs $\mathcal{G}_1, \mathcal{G}_2 \in [\mathcal{G}]$, with $\mathcal{G}_1$ having the feature, and $\mathcal{G}_2$ not having the feature.*

In other words, there is not yet a characterization of the analogous concept to the *essential graph* for maximal ancestral graphs (see Andersson et al. (1997)).

In addition the object $\mathbf{\Phi}$ is not a graphical model in the following sense:

1. It includes other markings in addition to vertices and edges.
2. There is no graphical characterization of which objects $\mathbf{\Phi}^*$ containing the same symbols as a PAG are such that for some $\mathcal{G}$, $\mathbf{\Phi}^* = \mathbf{\Phi}(\mathcal{G})$. Such a characterization is important for (b) above.

Spirtes et al. (1993) do provide a global Markov property for these graphs, called *definite d-connection*. The purpose of the next section is to make steps towards such a characterization in the case of arrowheads.

2.2 Join Operation: Definitions and Notation

Here I define the join operation as a method of identifying the arrowheads common to a set of Markov equivalent ancestral graphs. By definition, a set of Markov equivalent maximal ancestral graphs are required to have the same vertex set and adjacencies. The join operation can be thought of as an AND operation on the "arrowheads" of the set of

Markov equivalent ancestral graphs being joined, and thus an OR operation on the "tails" of these graphs.

Definition 2.2.1 *Let $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ be graphs with the same adjacencies. A joined graph, $\mathcal{H}$ is any graph constructed in the following way:*

(i) $\mathcal{H}$ has the same adjacencies as $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$,

(ii) For all adjacent α and β, add an arrowhead at β on the $\{\alpha, \beta\}$ edge if and only if there is an arrowhead at β on the $\{\alpha, \beta\}$ edge in all $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$.

The symbol "$\vee$" will be used to denote joining such that $\mathcal{G}_1 \vee \mathcal{G}_2$ should be read "$\mathcal{G}_1$ join $\mathcal{G}_2$".

In general I will let $\mathcal{H}$ refer to a joined graph formed by joining any number of Markov equivalent maximal ancestral graphs. We will also generically refer to these maximal ancestral graphs as $\mathcal{G}$. Richardson and Spirtes (2000) showed that for every non-maximal ancestral graph $\mathcal{G}$, there exists a unique maximal ancestral graph which is formed by adding appropriate bi-directed ($\leftrightarrow$) edges to $\mathcal{G}$ (see Figure 2.3). Hence I restrict my attention to joining sets of Markov equivalent maximal ancestral graphs in the remainder of this paper.

Proposition 2.2.1 *The join operation is commutative and associative.*

Proof:

This result is immediate by the definition of the join operation when viewed as an AND operation on the "arrowheads" and an OR operation on the "tails" of the set of Markov equivalent ancestral graphs being joined (since the AND and OR operations are commutative and associative). $\square$

Lemma 2.2.1 shows that the joined graph $\mathcal{H}$ is well-defined: if the join operation was not commutative or associative, then $\mathcal{H}$ would not be unique, and the resulting graph would be different depending on the order in which the ancestral graphs were joined.

Figure 2.7 provides an example of a joined graph. Note that since there are arrowheads that meet the undirected edge $x - w$ in the joined graph, $\mathcal{H}$ is not ancestral as it violates

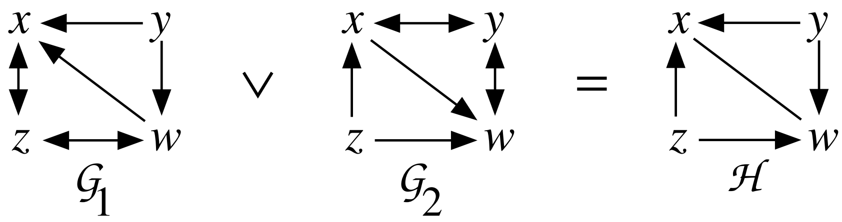


Figure 2.7: An example of joining two Markov equivalent ancestral graphs in which the joined graph is not ancestral.

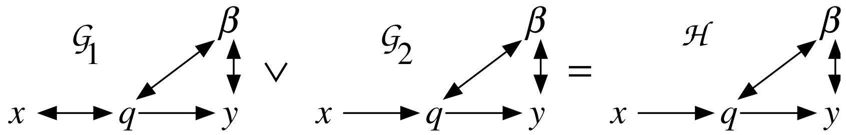


Figure 2.8: An example of joining two Markov equivalent ancestral graphs in which the joined graph is itself a member of the equivalence class.

condition *(c)* of Definition 2.1.1. Figure 2.8 shows another example of two Markov equivalent graphs being joined. Here, $\mathcal{H}$ is itself a member of the equivalence class of ancestral graphs.

I use the following meta-symbol for endpoints in either an ancestral graph or a joined graph:

1. "$\alpha - ?\beta$" is used to denote that there is a tail at α in the graph, on the edge between α and β, and that there may be a tail or an arrowhead at the β end of this edge.
2. "$\alpha \leftarrow ?\beta$" is used to denote that there is an arrowhead at α, and either an arrowhead or a tail at β on the edge between α and β.
3. "$\alpha ? - ?\beta$" is used to denote that there could be an arrowhead or tail at either end of the $\langle\alpha, \beta\rangle$ edge.

Note that the above notation is merely a shorthand since I only consider graphs with edges that are directed, bi-directed or undirected.

By joining maximal ancestral graphs as outlined in Definition 2.2.1, the resulting joined graph $\mathcal{H}$ is not ancestral in general, see Figure 2.7. This raises the question as to what is the appropriate Markov property for joined graphs? Ideally, I would wish $\mathcal{G}_1 \vee \mathcal{G}_2$ to be Markov equivalent to $\mathcal{G}_1$ and $\mathcal{G}_2$ if $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent. Here I define a *j-connecting* path for joined graphs.

Definition 2.2.2 *A path between α and β in a joined graph $\mathcal{H}$ is said to be "j-connecting given a set Z" (Z disjoint from $\{\alpha, \beta\}$ and possibly empty) if:*

(i) Every non-collider ($? - \gamma - ?, ? \to \gamma \to, \leftarrow \gamma \leftarrow ?$) on the path is not in Z,

(ii) Every collider ($? \to \gamma \leftarrow ?$) on the path is an ancestor of Z, and

(iii) No arrowheads meet undirected edges ($? \to \gamma -$).

If there is no path that j-connects α and β given Z, then α and β are "j-separated given Z".

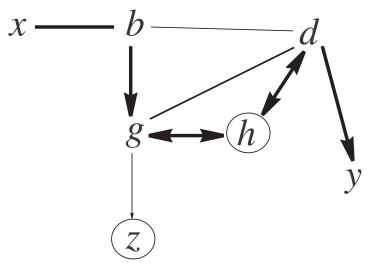


Figure 2.9: An example of a j-connecting path in a joined graph: x and y are j-connecting given $Z = \{z, h\}$.

Note that this definition is a natural extension of m-connection for ancestral graphs (and Pearl's d-connection for DAGs), with the qualifier that undirected edges meeting arrowheads form neither colliders nor non-colliders and a path containing such a vertex is never j-connecting. If we look back at the joined graph shown in Figure 2.7, we see that via j-connection, $\mathcal{H}$ encodes the same set of independence relations that the two ancestral graphs that gave rise to $\mathcal{H}$ encode, namely $y \perp\!\!\!\perp z$, because there are no j-connecting paths between

y and z in $\mathcal{H}$ (the path $z \to x - w \leftarrow y$ is not j-connecting). Figure 2.9 shows another example of a j-connecting path. Here, some vertices in Z are descendants of colliders on the path between x and y.

The definitions of discriminating paths and inducing paths for joined graphs remain the same as for ancestral graphs. Here I extend the concept of maximality to joined graphs and in Section 2.6.1 I show that the graph $\mathcal{H}$ formed by joining Markov equivalent maximal ancestral graphs is itself maximal.

Definition 2.2.3 *A joined graph $\mathcal{H}$ is said to be "maximal" if, for every pair of non-adjacent vertices α, β there exists a set $Z(\alpha, \beta \notin Z)$, such that α and β are j-separated conditional on Z.*

The concept of maximality for joined graphs is analogous to that for ancestral graphs in that a maximal joined graph is a joined graph, $\mathcal{H}$, such that no more edges can be added to $\mathcal{H}$ without changing the set of independence relations encoded by $\mathcal{H}$ via j-separation.

2.3 Joining an Entire Equivalence Class: sup$[\mathcal{G}]$

Let us make the following definition:

$$\sup[\mathcal{G}] = \vee_{\mathcal{G}' \in [\mathcal{G}]} \mathcal{G}'$$

$[\mathcal{G}]$ represents the equivalence class containing the ancestral graph $\mathcal{G}$. In other words, sup$[\mathcal{G}]$ is the analogy of the essential graph for Markov equivalent maximal ancestral graphs. One way of constructing sup$[\mathcal{G}]$ would be to determine every member of an equivalence class of ancestral graphs and then to join each one in turn until the entire class has been joined (See Figure 2.10).

From another perspective, consider the diagram shown in Figure 2.11. There is the space of maximal ancestral graphs (DAGs are contained in this space), which is a subspace of the set of (extended) joined graphs. In general, sup$[\mathcal{G}]$ lives in the set of extended joined graphs space. It has been shown by Spirtes and Richardson (1997) which MAGs are mapped into the same independence model under m-separation. Here I prove that the set of joined

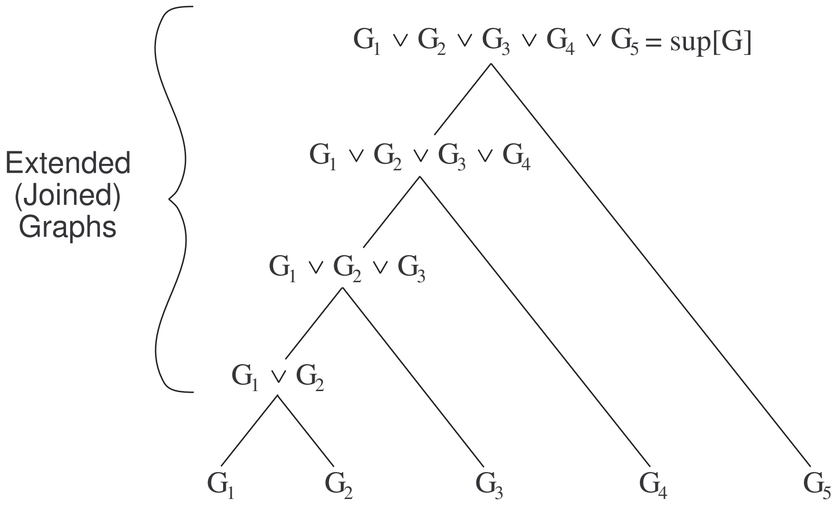


Figure 2.10: Schematic of Joining an Entire Equivalence Class of Ancestral Graphs to Construct sup$[\mathcal{G}]$.

Markov equivalent maximal ancestral graphs formed by joining members of an equivalence class map into the same independence model via j-connection. Of central concern is that sup$[\mathcal{G}]$ maps into the same independence model as do the ancestral graphs that gave rise to sup$[\mathcal{G}]$, but if the (extended) joined graph $\mathcal{H}$, formed by joining a subset of the graphs in $[\mathcal{G}]$, can be shown to be Markov equivalent to the set of Markov equivalent graphs that gave rise to $\mathcal{H}$, then the Markov equivalence of sup$[\mathcal{G}]$ to every graph $\mathcal{G}$ trivially follows.

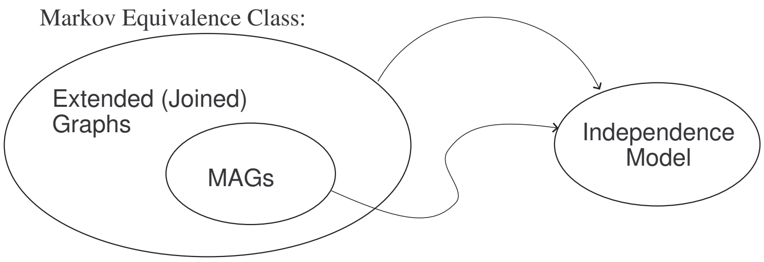


Figure 2.11: Schematic of Independence Model for Joined Graphs. See text for further details.

The joined graphs $\mathcal{H}$ shown in Figures 2.7 and 2.8 are examples of sup[$\mathcal{G}$]. Figure 2.12 shows three more examples of $\mathcal{G}$ and sup[$\mathcal{G}$]. In general, sup[$\mathcal{G}$] may be an undirected graph, a DAG, an ancestral graph or none of the above (see Figure 2.7). It is left to the reader to verify that the sup[$\mathcal{G}$] takes the form shown.

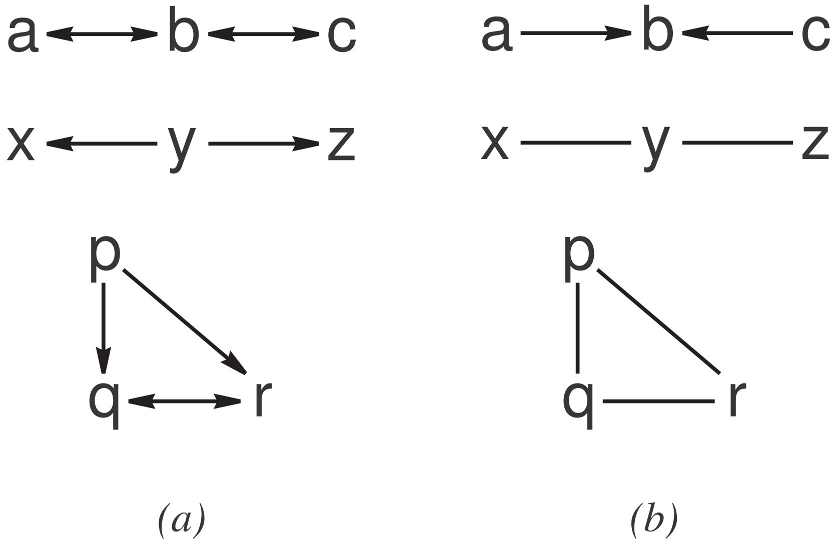


Figure 2.12: Column (a) maximal ancestral graphs $\mathcal{G}$; (b) the corresponding graphs sup[$\mathcal{G}$]

Figure 2.13 show all joined graphs that represent the Markov equivalence classes of maximal ancestral graphs over four vertices. The number of distinct classes given by permutation of the vertices is provided in parentheses. There are 251 classes in total: 185 are DAGS (left-hand column) because they contain at least one ancestral graph that contains no bi-directed or undirected edges; 3 do not contain any DAGs, but do contain the corresponding undirected graph (centre column); and 63 not contain any DAGs or undirected graphs (right-hand column).

2.4 Towards Characterizing Joined Graphs

To date, no full characterization of graphs formed by joining Markov equivalent maximal ancestral graphs is readily available. This section presents structural inferences that can

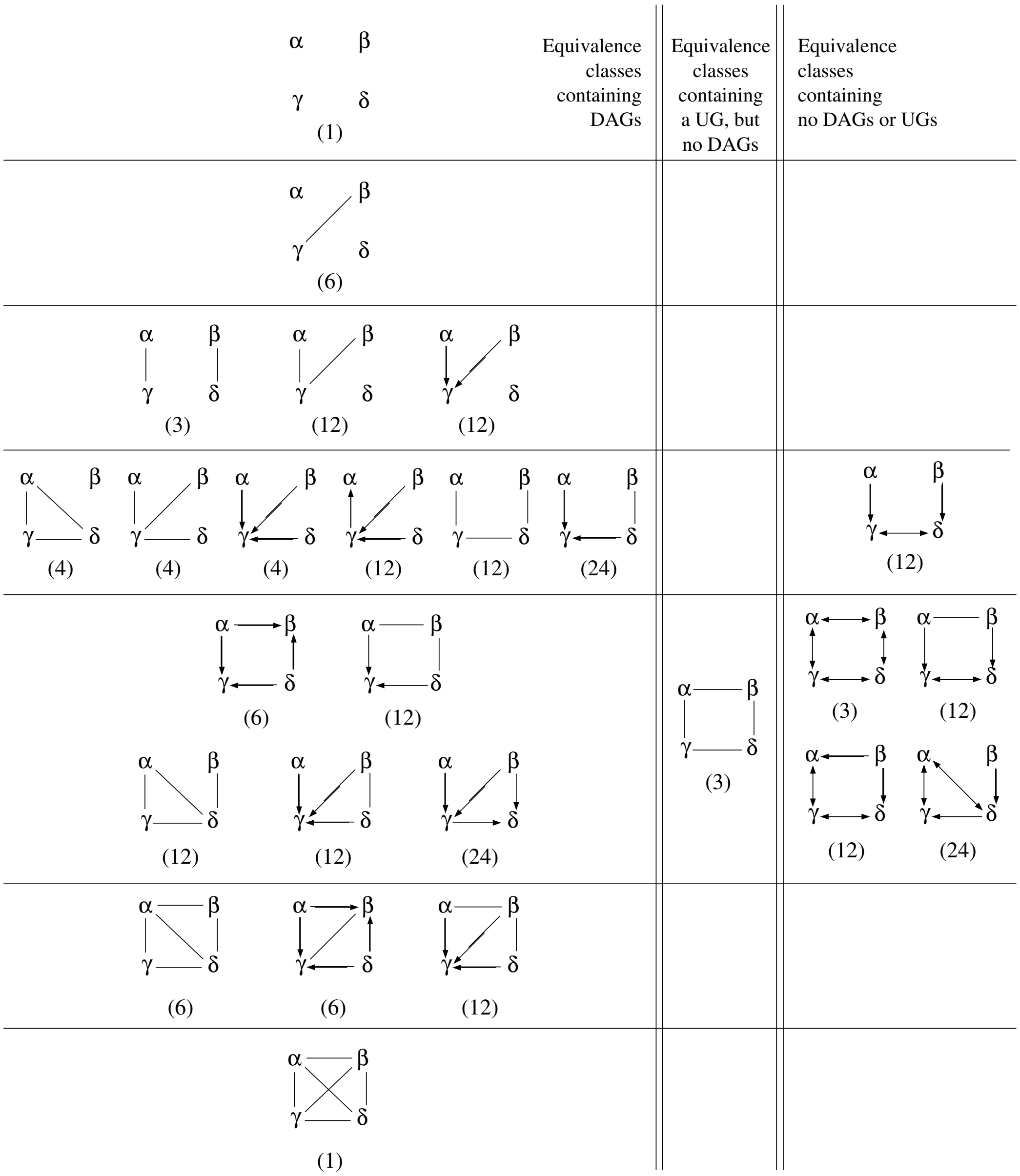


Figure 2.13: Joined graphs representing all Markov equivalent maximal ancestral graphs on four variables. See text for further details.

be made about joined graphs. For instance, as with ancestral graphs, the configurations "$\rightarrow \gamma -$" and "$\leftrightarrow \gamma -$" do not occur as induced sub-graphs in joined graphs. I also conjecture that the graph resulting from joining an entire equivalence class of ancestral graphs can be more constrained than that obtained by joining only a few members of an equivalence class. The following few lemmas start to examine the structure of joined graphs in general.

Lemma 2.4.1 *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent ancestral graphs, then $\mathcal{G}_1 \vee \mathcal{G}_2$ will retain the unshielded colliders in $\mathcal{G}_1$ and $\mathcal{G}_2$*

Proof:

Suppose that y is an unshielded collider in $\mathcal{G}_1$, and is adjacent to x and z. Then $x? \rightarrow y \leftarrow ?z$ is in $\mathcal{G}_1$. By Theorem 2.1.1, $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders and $x? \rightarrow y \leftarrow ?z$ is in $\mathcal{G}_2$. Since there is an arrowhead at y on the (x, y) edge in $\mathcal{G}_1$ and $\mathcal{G}_2$, then by definition, $\mathcal{G}_1 \vee \mathcal{G}_2$ also has an arrowhead at y on this edge. Similarly, there is an arrowhead at y on the (y, z) edge in $\mathcal{G}_1$ and $\mathcal{G}_2$, so there must be an arrowhead at y on this edge in the joined graph. □

Corollary 2.4.1 *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent ancestral graphs, then $\mathcal{G}_1 \vee \mathcal{G}_2$ will not contain any chordless partially directed cycles.*

The graphs shown in Figure 2.7 are not *chordless* because there is an edge between x and w. If however, there is a path in a graph $\pi = \langle x, y, w, z, x \rangle$ with neither $\{x, w\}$ nor $\{z, y\}$ being adjacencies in the graph, then π is a chordless four-cycle.

Proof: (Corollary 2.4.1)

If any four vertices, say $\{a, b, c, d\}$, form a chordless partially directed cycle in $\mathcal{G}_1 \vee \mathcal{G}_2$, then these vertices also form a chordless cycle in $\mathcal{G}_1$ and $\mathcal{G}_2$ since all three graphs have the same adjacencies. If there were no colliders, then either $\langle a, b, c, d \rangle$ would be a directed cycle, or there would be a configuration of an arrow pointing to an undirected edge (i.e. a flag). In $\mathcal{G}_1$, there is at least one collider in the subgraph formed by these vertices, and this collider is unshielded (see Figure 2.14). In either case, $\mathcal{G}_1$ would not be ancestral. By Theorem 2.1.1 $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same unshielded colliders. Since $\langle a, b, c, d \rangle$ must contain at least one collider in $\mathcal{G}_1$, then that vertex must also be a collider in the subgraph formed by $\{a, b, c, d\}$ in $\mathcal{G}_2$. By Corollary 2.4.1 $\mathcal{G}_1 \vee \mathcal{G}_2$ will retain any unshielded colliders, so the subgraph $\{a, b, c, d\}$ in $\mathcal{G}_1 \vee \mathcal{G}_2$ also contains colliders. Thus $\mathcal{G}_1 \vee \mathcal{G}_2$ does not include chordless partially directed four-cycles. The above argument can easily be extended to include chordless partially directed cycles with more than four vertices. □

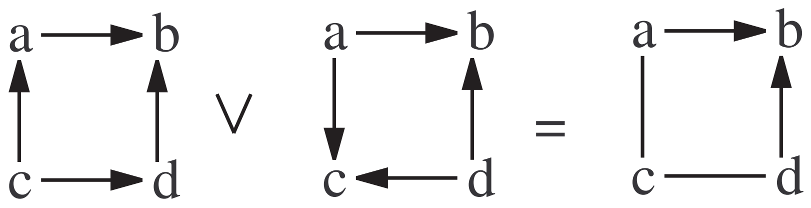


Figure 2.14: Example of a chordless partially directed cycle in a graph formed by joining two maximal ancestral graphs that are not Markov equivalent. See text for further details.

Lemma 2.4.2 *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent ancestral graphs, then $\mathcal{G}_1 \vee \mathcal{G}_2$ may contain partially directed cycles.*

Proof:

See Figure 2.15 for an example of joining two Markov equivalent ancestral graphs that gives rise to a partially directed cycle. □

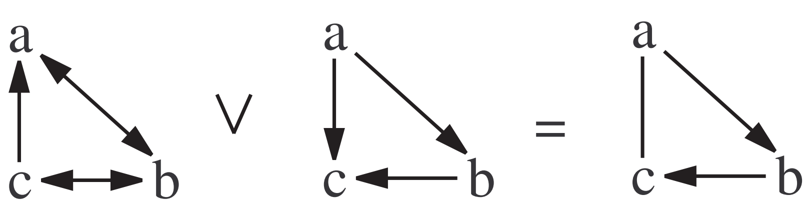


Figure 2.15: Example of a partially directed cycle in a joined graph. See text for further details.

Lemma 2.4.3 *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent ancestral graphs, then the configuration $a \rightarrow x - y \leftarrow b$, with the respective pairs of vertices $(a, y), (b, x)$ not adjacent cannot occur in $\mathcal{G}_1 \vee \mathcal{G}_2$.*

Proof:

Note that the configuration $a \rightarrow x - y \leftarrow b$ is not ancestral. Therefore the (x, y) edge is directed in $\mathcal{G}_1$ and in $\mathcal{G}_2$. Suppose that $x \rightarrow y$ in $\mathcal{G}_1$. Then $x \rightarrow y \leftarrow b$ forms an unshielded collider and $x \rightarrow y$ is in $\mathcal{G}_1$ by Lemma 2.4.1, and thus in $\mathcal{G}_1 \vee \mathcal{G}_2$, which is a contradiction. A similar argument holds if we consider $x \leftarrow y$ in $\mathcal{G}_1$. $\square$

2.5 Real Edges in Joined Graphs

If an edge is oriented the same way in all graphs $\mathcal{G}$ that were joined to form $\mathcal{H}$, then that edge is said to be "*real*" in $\mathcal{H}$. By virtue of the join operation, it is possible to infer the presence of arrowheads and tails in joined graphs under certain circumstances. The following lemmas describe some of these situations.

Lemma 2.5.1 *All bi-directed edges in a joined graph are real. Furthermore, if $\alpha ? \rightarrow \beta - \gamma$ is a sub-graph of $\mathcal{H}$, and α and β are adjacent, then the $\beta - \gamma$ edge is not real.*

Proof:

The first part is trivial since by definition of the join operation, an arrowhead appears at a vertex in the joined graph $\mathcal{H}$ if and only if there is an arrowhead at that vertex in all ancestral graphs that gave rise to $\mathcal{H}$. Also, no ancestral graph contains undirected edges meeting arrowheads, so the only way that an undirected edge could meet an arrowhead in a joined graph (using the example given in the lemma) is if there is at least one ancestral graph that gave rise to $\mathcal{H}$ with a tail at β, i.e. $\alpha \rightarrow \beta \rightarrow \gamma$, since otherwise the graph would not be ancestral, and at least one graph that contains $\alpha \rightarrow \beta \leftarrow \gamma$. $\square$

Lemma 2.5.2 *In a joined graph $\mathcal{H}$, formed by joining maximal ancestral graphs, if $\gamma \rightarrow \beta ? - ? \delta ? \rightarrow \gamma$ occurs and $\gamma \rightarrow \beta$ is real, then $\beta \leftarrow ? \delta$ also occurs in $\mathcal{H}$.*

Proof:

Note that $\gamma \to \beta — ?\delta ? \to \gamma$ is not ancestral itself (under any substitution of the question marks by either an arrowhead or tail). In any $\mathcal{G}$ that gave rise to $\mathcal{H}$, γ is not an ancestor of δ. See Figure 2.16. Thus, since by hypothesis $\gamma \to \beta$ in any graph $\mathcal{G}$ joined to form $\mathcal{H}$, it follows that $\beta \leftarrow ?\gamma$ and hence $\beta \leftarrow ?\delta$ in $\mathcal{H}$. □

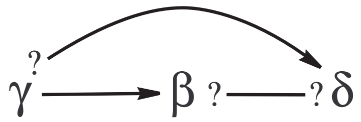


Figure 2.16: Diagram for proof of Lemma 2.5.2. See text for further details.

Lemma 2.5.3 *In a joined graph $\mathcal{H}$, formed by joining ancestral graphs, if $\gamma \to \beta \to \delta ? — ?\gamma$ occurs and either $\gamma \to \beta$ is real or $\beta \to \delta$ is real, then $\gamma \to \delta$ also occurs in $\mathcal{H}$. Furthermore, if both $\gamma \to \beta$ and $\beta \to \delta$ are real, then $\gamma \to \delta$ is real too.*

Proof:

First consider the case in which the $\gamma \to \beta$ edge is real (see Figure 2.17). Then, β is not an ancestor of γ in any $\mathcal{G}$ that gave rise to $\mathcal{H}$. If in $\mathcal{H}$ the $\{\gamma, \delta\}$ edge is undirected, or there is an arrowhead at γ on this edge, then there is some $\mathcal{G}$ that gave rise to $\mathcal{H}$ that is not ancestral. So, $\gamma \to \delta$ is in $\mathcal{H}$. A similar argument holds for the case in which the $\beta \to \delta$ edge is real.

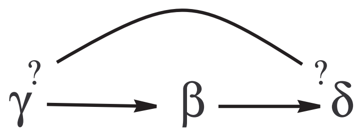


Figure 2.17: Diagram for proof of Lemma 2.5.3. See text for further details.

If both edges $\gamma \to \beta$ and $\beta \to \delta$ are real, then $\gamma \to \delta$ also occurs in $\mathcal{H}$ and this edge is real because otherwise there is some $\mathcal{G}$ that gave rise to $\mathcal{H}$ that is not ancestral. $\square$

2.5.1 Inferring Discriminating Paths

The following lemma and corollary allow us to infer the presence of discriminating paths.

Lemma 2.5.4 *Suppose $\mathcal{H}$ is a graph formed by joining a set of Markov equivalent maximal ancestral graphs. If there is a discriminating path in $\mathcal{H}$ then this discriminating path is present in every $\mathcal{G}$ joined to form $\mathcal{H}$.*

Proof:
Suppose in some joined graph $\mathcal{H}$ there is a path U as described in Definition 2.1.8. Label the colliders on the path between x and B as $q_1, q_2, \ldots, q_p$, such that q_1 is adjacent to x, and q_p is adjacent to B, as shown in Figure 2.18. Note that $\langle x, q_1, q_2, \ldots, q_p, B\rangle$ forms a collider path in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ because all arrowheads in $\mathcal{H}$ are also present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Recall that x and y are not adjacent. There is an unshielded non-collider at q_1 on the path $\langle x, q_1, y\rangle$, but $x? \to q_1$. Because all $\mathcal{G}$ that gave rise to $\mathcal{H}$ are Markov equivalent, by Theorem 2.1.1 q_1 is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. I will now show by induction that all $q_m, 2 \leq m \leq p$ are also parents of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

For $m = 2$, $\langle x, q_1, q_2, y\rangle$ discriminates q_2 to be a non-collider in $\mathcal{H}$. Since q_1 is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, this discriminating path is present in all such $\mathcal{G}$, and q_2 is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Assume for $m < p$ that $\langle x, q_1, q_2, \ldots, q_{m-1}, q_m, y\rangle$ discriminates q_m to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ so that q_m is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Then, $\langle x, q_1, q_2, \ldots, q_m, q_{m+1}, y\rangle$ discriminates $\langle q_m, q_{m+1}, y\rangle$ to be a non-collider in $\mathcal{H}$. Because $< q_1, q_2, \ldots, q_m >$ are parents of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, this discriminating path is present in all such $\mathcal{G}$ so $\langle q_m, q_{m+1}, y\rangle$ forms a non-collider in all such $\mathcal{G}$. Since $q_m? \to q_{m+1}$, $q_{m+1} \to y$ in $\mathcal{H}$, q_{m+1} is also a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. So, by induction, $< q_1, q_2, \ldots, q_p >$ are all parents of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

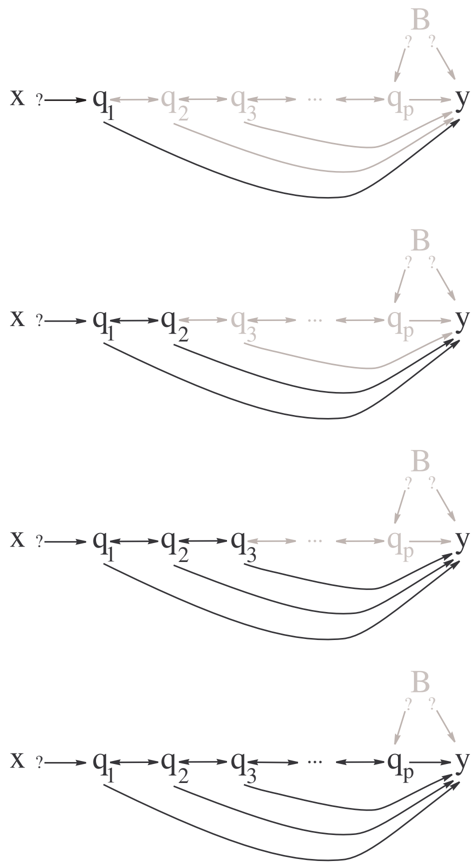


Figure 2.18: Discriminating Path for Proof of Lemma 2.5.4.

But then $\langle x, q_1, q_2, \ldots, q_p, B, y\rangle$ (i.e. U) forms a discriminating path for B that is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. $\square$

Corollary 2.5.1 *If a collider path $q = \langle q_1, \ldots, q_p\rangle$ is present in all ancestral graphs that gave rise to the joined graph $\mathcal{H}$, and $\mathcal{U} = \langle x, q_1, \ldots, q_p, B, y\rangle$ is a discriminating path for B in some $\mathcal{G}$ that gave rise to $\mathcal{H}$, then $\mathcal{U}$ is also a discriminating path for B in $\mathcal{H}$.*

Proof:

Since the collider path q is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, it is also present in $\mathcal{H}$. So it suffices to show that each $q_i, 1 \leq i \leq p$ is a parent of y in $\mathcal{H}$.

Note that since there is a tail at q_i on the edges $q_i \to y, 1 \leq i \leq p$ in $\mathcal{G}$, there is a tail at q_i on the edges $q_i - ?y$ in $\mathcal{H}$. Furthermore, since x and y are not adjacent in $\mathcal{G}$ or $\mathcal{H}$, $\langle x, q_1, y \rangle$ forms an unshielded non-collider in all $\mathcal{G}$, and there is an edge $x? \to q_1$ in all $\mathcal{G}$, it follows that q_1 is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. But then $\langle x, q_1, q_2, y \rangle$ forms a discriminating path for q_2 in $\mathcal{H}$, and by Lemma 2.5.4, $\langle q_1, q_2, q_3 \rangle$ is discriminated to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Assume for $m < p$ that $\langle x, q_1, q_2, \ldots, q_m, y \rangle$ discriminates $\langle q_{m-1}, q_m, y \rangle$ to be a non-collider in $\mathcal{H}$. By Lemma 2.5.4, $\langle q_{m-1}, q_m, y \rangle$ is also discriminated to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and so q_m is a parent of y in $\mathcal{H}$. But then $\langle x, q_1, q_2, \ldots, q_m, q_{m+1}, y \rangle$ discriminates $\langle q_m, q_{m+1}, y \rangle$ to be a non-collider in $\mathcal{H}$ and by Lemma 2.5.4 $\langle q_m, q_{m+1}, y \rangle$ is a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Thus, since $q_m? \to q_{m+1}$, q_{m+1} is a parent of y in $\mathcal{H}$. By induction, $\langle q_1, q_2, \ldots, q_p \rangle$ are all parents of y in $\mathcal{H}$. Now it is easy to see that $\langle x, q_1, \ldots, q_p, B, y \rangle$ forms a discriminating path for B in $\mathcal{H}$.

□

Just as for ancestral graphs, there are notions of inducing paths and maximal graphs for joined graphs:

Definition 2.5.1 *An "inducing path" for a joined graph $\mathcal{H}$ is a collider path with distinct vertices such that every interior vertex of the path is an ancestor of at least one endpoint.*

Definition 2.5.2 *A joined graph $\mathcal{H}$ is said to be "maximal" if there is no pair of non-adjacent vertices $< \alpha, \beta >$ in $\mathcal{H}$ such that there is at least one inducing path between α and β.*

2.5.2 Collider Paths in Joined Graphs

Since both discriminating and inducing paths are types of collider paths, it is fruitful to look at properties of collider paths, in general. The following lemma and corollaries examine which non-consecutive vertices along a *minimal* collider path may be adjacent, and whether the orientation of such an edge can be inferred, where:

Definition 2.5.3 *A collider path is "minimal" if no subsequence of the vertices on the path form a shorter collider path with the same endpoints.*

The collider sub-path of a minimal j-connecting path is a minimal collider path, as is a minimal inducing path.

Definition 2.5.4 *Let $\mu = \langle \mu_0, \mu_1, \ldots, \mu_n \rangle$ be a path in a graph. Any vertex μ_j is non-consecutive to μ_i, $1 \leq i \leq n$ if $|j-i| > 1$, provided such a μ_j exists.*

Lemma 2.5.5 *Let $\mathcal{H}$ be a graph formed by joining any number of Markov equivalent maximal ancestral graphs. Suppose there is a minimal collider path μ between vertices μ_0 and μ_n such that μ_0 and μ_n are not adjacent in the joined graph $\mathcal{H}$. Let $\mu_1, \mu_2, \ldots, \mu_{n-1}$ be the interior vertices along this path (i.e. the non-endpoints). If μ_i and μ_j are non-consecutive vertices along μ and $\mu_i - ?\mu_j$ in $\mathcal{H}$, then $\mu_i \to \mu_j$ in all ancestral $\mathcal{G}$ that gave rise to $\mathcal{H}$. So, $\mu_i \to \mu_j$ in $\mathcal{H}$ and this edge is real.*

Proof:

Suppose, for a contradiction, that μ is such that there is an edge $\mu_i - ?\mu_j$ in $\mathcal{H}$, and the tail at μ_i is not real.

Let $\{\mu_i, \mu_j\}$ be such that there is no k, m where $|m-k| > |j-i|$ such that there is an edge $\mu_k - ?\mu_m$ in $\mathcal{H}$ that is not present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, i.e. $|j-i|$ is as large as possible. (I will refer to this property as "the construction of i and j"). Suppose, without loss of generality, that $i < j$. Note that $\mu_k \leftrightarrow \mu_m, |k-m| > 1$ cannot occur in $\mathcal{H}$ because otherwise μ is not minimal. Furthermore, the distance between i and j cannot be greater than $n-1$ since μ_0 and μ_n are not adjacent.

1. Claim: (μ_{i-1}, μ_j) are adjacent.

   If (μ_{i-1}, μ_j) are not adjacent, $(\mu_{i-1}, \mu_i, \mu_j)$ forms an unshielded non-collider, which by Lemma 2.5.4 implies that all $\mathcal{G}$ that gave rise to $\mathcal{H}$ also have an unshielded non-collider at μ_i. Since ancestral graphs do not contain undirected edges that meet arrowheads and the tail at μ_i is real, $\mu_i \to \mu_j$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ (i.e. $\mu_i \to \mu_j$ should occur in $\mathcal{H}$). But this is a contradiction because we are assuming that there is at least one graph $\mathcal{G}$, that gave rise to $\mathcal{H}$, without $\mu_i \to \mu_j$. So, (μ_{i-1}, μ_j) are adjacent in $\mathcal{H}$. Recall that $\mu_{i-1} \leftrightarrow \mu_j$ cannot occur in $\mathcal{H}$ because otherwise μ is not minimal.

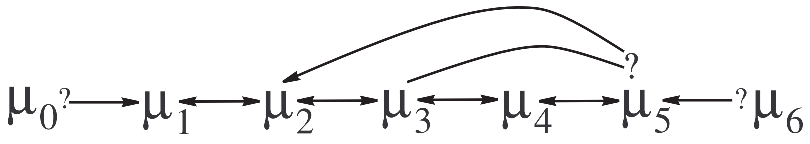


Figure 2.19: Collider Path for Proof of Lemma 2.5.5 supposing $\mu_{i-1}? - \mu_j$ in $\mathcal{H}$ where $i = 3, j = 5$.

2. Suppose $\mu_{i-1}? - \mu_j$ in $\mathcal{H}$ (see Figure 2.19).

   Then, by the construction of i and j, $\mu_{i-1} \leftarrow \mu_j$ in $\mathcal{H}$ and this edge is real since $| (i-1) - j | > | i - j |$. But by Lemma 2.5.3 we reach a contradiction: there is a tail at μ_i in $\mathcal{H}$ on the $\mu_i - ?\mu_j$ edge, which implies that there is at least one $\mathcal{G}$ that gave rise to $\mathcal{H}$ that also has a tail at μ_i on this edge. This graph would not be ancestral because it would contain the following sub-graph: $\mu_{i-1} \leftrightarrow \mu_i \rightarrow \mu_j \rightarrow \mu_{i-1}$.
3. Suppose $\mu_{i-1} \rightarrow \mu_j$ in $\mathcal{H}$ (see Figure 2.20). Again, this edge is real (by construction of i and j).

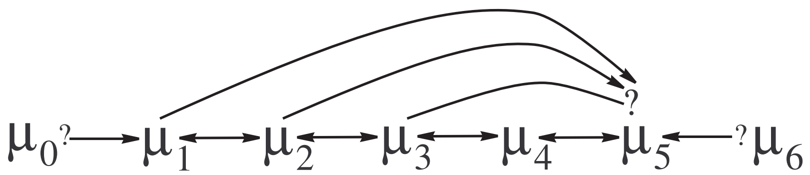


Figure 2.20: Collider Path for Proof of Lemma 2.5.5 supposing $\mu_{i-1} \rightarrow \mu_j$ in $\mathcal{H}$.

If $i > 1$ and μ_{i-2} is not adjacent to μ_j then $\langle \mu_{i-2}, \mu_{i-1}, \mu_i, \mu_j \rangle$ forms a discriminating path for μ_i in $\mathcal{H}$. Here we reach a contradiction: there is at least one $\mathcal{G}$ that contains the edge $\mu_i \rightarrow \mu_j$. In this graph, $\langle \mu_{i-1}, \mu_i, \mu_j \rangle$ is discriminated to be a non-collider, so by Lemma 2.5.4 any Markov equivalent ancestral graph must also have a non-collider at μ_i which would imply that the $\mu_i \rightarrow \mu_j$ edge in $\mathcal{H}$ was real.

If for all $k < i$, $\mu_k? \rightarrow \mu_j$ then $\mu_0? \rightarrow \mu_j$ and the collider path μ is not minimal. Let μ_k be the largest k such that $k < i$ and there is no edge $\mu_k? \rightarrow \mu_j$ in $\mathcal{H}$. (It follows from Case 2. that $k < (i-1)$). Then for all r, $k < r < i$, $\mu_r? \rightarrow \mu_j$ in $\mathcal{H}$. This implies that for all r, $k < r < i$, $\mu_r \to \mu_j$ in $\mathcal{H}$ by the minimality of the path μ. And now we see that for all r, $k < r < i$, $\mu_r \to \mu_j$ in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ by the construction of i and j. There are two cases to be considered:

i) If μ_k and μ_j are not adjacent then $< \mu_k, \mu_{k+1}, \ldots, \mu_{i-1}, \mu_i, \mu_j >$ forms a discriminating path for μ_i in all $\mathcal{G}$. By Lemma 2.5.4, $\langle \mu_{i-1}, \mu_i, \mu_j \rangle$ forms a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ and hence $\mu_i \to \mu_j$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. But this is a contradiction because we are assuming that the $\mu_i \to \mu_j$ edge in $\mathcal{H}$ is not real.

ii) If μ_k and μ_j are adjacent then $\mu_k ?\!-\!\mu_j$ in $\mathcal{H}$ (by definition of k) which leads to a contradiction: If $\mu_k ?\!-\!\mu_j$ in $\mathcal{H}$ then $\mu_k ?\!\to \mu_{k+1} \to \mu_j -\!? \mu_k$ in $\mathcal{H}$. Since $k < i - 1$, the μ_{k+1} edge is real so $\mu_k ?\!\to \mu_{k+1} \to \mu_j \to \mu_k$ in at least one $\mathcal{G}$ that was joined to form $\mathcal{H}$, and such a $\mathcal{G}$ would violate the ancestral condition. $\square$

Corollary 2.5.2 *Let μ be a minimal collider path of length n in the joined graph $\mathcal{H}$. If any $\mu_r, 1 \le r \le (n-1)$ is adjacent to an endpoint that does not occur directly before or after μ_r along the path μ (i.e. excluding $\langle \mu_0, \mu_1 \rangle$ and $\langle \mu_{n-1}, \mu_n \rangle$), then μ_r is a parent of the endpoint in $\mathcal{H}$ and this edge is real.*

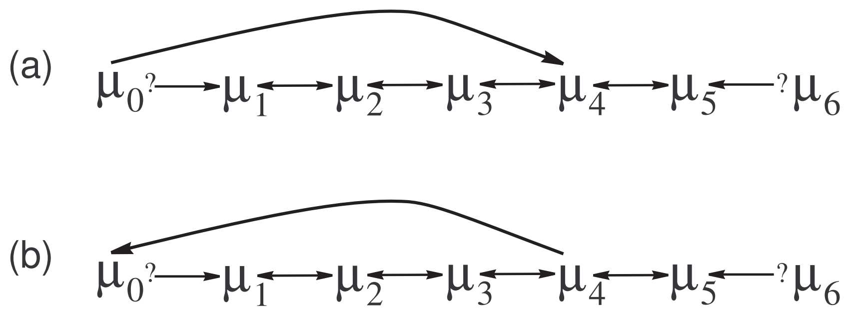


Figure 2.21: Diagram for Proof of Lemma 2.5.2. See text for details.

Proof:

By Lemma 2.5.5 there is an directed edge between μ_r and the endpoint. If the endpoint is a parent of μ_r then the minimality of the path is violated, as shown in Figure 2.21(a).

Therefore μ_r is a parent of the endpoint, and by Lemma 2.5.5, this edge is real, as shown in Figure 2.21(b). □

Corollary 2.5.3 *Let $\mathcal{H}$ be a graph formed by joining any number of Markov equivalent maximal ancestral graphs. Suppose there is a minimal collider path μ between vertices μ_0 and μ_n such that μ_0 and μ_n are not adjacent in the joined graph $\mathcal{H}$. Let $\mu_1, \mu_2, \ldots, \mu_{n-1}$ be the interior vertices along this path (i.e. the non-endpoints). Furthermore, suppose that $\mu_i \rightarrow \mu_j$, $|j-i| > 1$ occurs in $\mathcal{H}$ for some i, $1 < i, j < n$. Suppose, without loss of generality, that $i < j$. If μ_i is adjacent to μ_{j+1} or μ_{j-1} (provided these vertices exist), then these edges are oriented out of μ_i in $\mathcal{H}$. Similarly, if μ_j is adjacent to μ_{i-1} or μ_{i+1} (provided these vertices exist), then these edges is oriented into μ_j in $\mathcal{H}$. Furthermore, these edges are oriented in the above fashion in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and hence are real.*

Proof:

From Lemma 2.5.5 we know that the $\mu_i \rightarrow \mu_j$ edge in $\mathcal{H}$ is real. Furthermore, we know that there are no bi-directed edges between non-consecutive vertices on μ in $\mathcal{H}$ because μ is minimal. Suppose μ_i is adjacent to μ_{j+1} such that $\mu_i \leftarrow \mu_{j+1}$ in $\mathcal{H}$. Then there is at least one ancestral $\mathcal{G}$ that gave rise to $\mathcal{H}$ in which $\mu_i \leftarrow \mu_{j+1}$ occurs. If μ_{j+1} is not an endpoint, then a contradiction because $\mu_{i-1} \leftrightarrow \mu_i \leftarrow \mu_{j+1} \rightarrow \mu_{i-1}$ would occur in this graph and this graph would not be ancestral. If μ_{j+1} or μ_{i-1} is an endpoint, then the path $\langle \mu_0, \ldots, \mu_i, \mu_n \rangle$ is a shorter collider path. An analogous contradiction would arise for the other cases outlined in the corollary. □

2.6 Inducing Paths in the Joined Graph

In the next section I prove that the graph formed by joining Markov equivalent maximal ancestral graphs is itself maximal. It is sufficient to only consider particular inducing paths between vertices, *minimal* inducing paths.

Definition 2.6.1 *Let μ be an inducing path in a joined graph with vertices $< \mu_0, \mu_2 \ldots, \mu_n >$, and let ψ_i be the number of edges on a shortest directed path between μ_i and an endpoint. Furthermore, let $\phi(\mu)$ be the total number of edges between the interior vertices and the* *endpoints on these paths, i.e.* $\phi(\mu) = \Sigma_{i=1}^{n-1}\psi_i$. *Then,* μ *is a "minimal inducing path" for vertices* μ_0 *and* μ_n *in a joined graph if:*

(i) there is no other inducing path between μ_0 *and* μ_n *with fewer vertices, and*

(ii) there is no inducing path μ' *with the same number of vertices as* μ *and* $\phi(\mu') < \phi(\mu)$.

It is easy to see that whenever there is an inducing path then there is a minimal inducing path. Consequently, if there is an inducing path with non-adjacent endpoints then there is a minimal inducing path with non-adjacent endpoints. I use the notion of a minimal inducing path to help infer the presence and/or orientation of edges between non-consecutive vertices along a collider path in a joined graph. Lemma 2.5.5 and Corollary 2.5.2 makes such inferences.

2.6.1 Outline of Proof of Maximality for Joined Graphs

To prove that joining Markov equivalent maximal ancestral graphs results in a maximal joined graph, I show that if there is a minimal inducing path, μ, in the joined graph, with endpoints x and y, then μ is present in all the ancestral graphs $\mathcal{G}$ that gave rise to the joined graph.

Since the graphs that are joined are assumed to be maximal this implies that x and y are adjacent in all $\mathcal{G}$, and hence in $\mathcal{H}$. Recall that inducing paths are collider paths such that each interior node is an ancestor of at least one endpoint. Consider any interior node, μ_i, $0 < i < n$ where n is the number of edges on the path μ, and let α_0 be the first descendant of μ_i on the directed path from μ_i to an endpoint. The crux of the proof lies in showing that μ_i is a parent of α_0 in all ancestral graphs $\mathcal{G}$ that gave rise to the joined graph.

Figure 2.22 can be thought of as a "road map" for proving Theorem 2.6.1.

Lemmas 2.6.2 to 2.6.5 combine to show that μ_i is a parent of α_0 in all ancestral graphs $\mathcal{G}$ that gave rise to the joined graph. Lemma 2.5.3 is an important result for proving Lemmas 2.6.2 to 2.6.5. It states that if non-consecutive vertices on a collider path are adjacent in the graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs, then

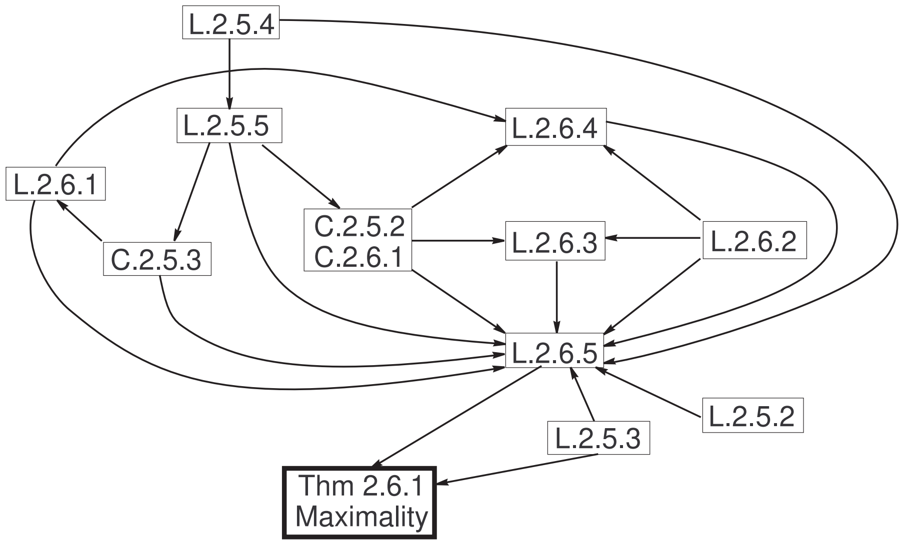


Figure 2.22: Road map for Proof of Pairwise Markov Property for Joined Graphs. See text for details.

that edge is directed, and has the same orientation in all ancestral graphs joined to give rise to $\mathcal{H}$. Lemma 2.6.1 states that non-consecutive vertices do not participate in triangular cycles with α_0. This result is important for proving the orientation of edges for which α_0 is an endpoint. Once it has been established that the $\mu_i \to \alpha_0$ edge is real, it is not so difficult to show that the ancestral path from μ_i to an endpoint in $\mathcal{H}$ is also ancestral in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. I now prove Lemma 2.6.1.

Lemma 2.6.1 *Let $\mathcal{H}$ be a graph resulting from joining any number of maximal ancestral graphs that are Markov equivalent to each other. Suppose μ is the shortest minimal inducing path in $\mathcal{H}$ with non-adjacent endpoints μ_0 and μ_n such that $n > 3$. Furthermore, let α_0 be a child of a vertex $\mu_j, 1 < j < (n-1)$. If such a path μ exists in $\mathcal{H}$, then the configurations $\mu_j \to \alpha_0 \to \mu_{j+1}$ and $\mu_{j-1} \leftarrow \alpha_0 \leftarrow \mu_j$ do not occur in $\mathcal{H}$ for $1 < j < (n-1)$.*

Note that the conditions of Lemma 2.6.1 require that μ_j is not an endpoint and does not succeed or precede an endpoint on μ. Thus, this lemma only applies to paths with $n > 4$. The basic argument of the proof of Lemma 2.6.1 is that if $\mu_j \to \alpha_0 \to \mu_{j+1}$ occurs in $\mathcal{H}$, then there are inducing paths on sub-paths of μ. Consequently, we can infer the presence of edges between non-consecutive edges along μ because of the minimality of the path μ. Arguing in this way, we can eventually infer an edge between μ_0 and μ_n, which is a contradiction.

Proof:

For a contradiction, suppose that $\mu_j \to \alpha_0 \to \mu_{j+1}$ occurs in $\mathcal{H}$.

If μ is of length greater than 3, then $\langle \mu_{j-1}, \mu_j, \mu_{j+1} \rangle$ forms an inducing path in $\mathcal{H}$ that is shorter than μ, so μ_{j-1} and μ_{j+1} are adjacent. By Lemma 2.5.5, the $\{\mu_{j-1}, \mu_{j+1}\}$ edge cannot be undirected. Furthermore, the $\{\mu_{j-1}, \mu_{j+1}\}$ edge cannot be bi-directed because then the minimality of the path is violated. There are two cases left to consider (see Figure 2.23):

a) $\mu_{j-1} \to \mu_{j+1}$ in $\mathcal{H}$

Since μ_{j-1} is not an endpoint $\langle \mu_{j-2}, \mu_{j-1}, \mu_j, \mu_{j+1} \rangle$ forms an inducing path in $\mathcal{H}$ that is shorter than μ, so μ_{j-2} and μ_{j+1} are adjacent. If μ_{j-2} is an endpoint then $\mu_{j-2} \to \mu_{j+1}$ by Corollary 2.5.2. If μ_{j-2} is not an endpoint then by Corollary 2.5.3, this edge is oriented $\mu_{j-2} \to \mu_{j+1}$ and by Lemma 2.5.5 this edge is real. By repeated applications of the shortest inducing path argument, $\langle \mu_1, \dots, \mu_{j-1}, \mu_j, \mu_{j+1} \rangle$ forms an inducing path that is shorter than μ, and each of $\mu_1, \dots, \mu_{j-1}$ are parents of μ_{j+1} and these edges are real (by Lemma 2.5.5). Since $j \neq n-1$, μ_0 is adjacent to μ_{j+1} because $\langle \mu_0, \mu_1, \dots, \mu_j, \mu_{j+1} \rangle$ forms an inducing path that is shorter than μ. By Corollary 2.5.3 $\mu_0 \to \mu_{j+1}$, but then we reach a contradiction because $\langle \mu_0, \mu_{j+1}, \mu_{j+2}, \dots, \mu_n \rangle$ is a shorter minimal inducing path between μ_0 and μ_n than μ.

b) $\mu_{j-1} \leftarrow \mu_{j+1}$ in $\mathcal{H}$

In this case, $\langle \mu_{j+2}, \mu_{j+1}, \mu_j, \mu_{j-1} \rangle$ forms an inducing path in $\mathcal{H}$ that is shorter than μ (because μ_{j+1} is an ancestor of μ_{j-1}, and μ_j is an ancestor of μ_{j+1} and thus an ancestor of μ_{j-1}). So μ_{j+2} and μ_{j-1} are adjacent. Furthermore, by Corollary 2.5.3, this edge is oriented $\mu_{j-1} \leftarrow \mu_{j+2}$. By repeated applications of the shortest inducing path argument, $\langle \mu_{n-1}, \dots . \mu_{j+1}, \mu_j, \mu_{j-1} \rangle$ forms an inducing path that is shorter than μ, and each of $\mu_{n-1}, \dots, \mu_{j+1}$ are parents of μ_{j-1}. Since $j \neq 1$, $\mu_n \to \mu_{j-1}$ because $\langle \mu_n, \mu_{n-1}, \dots, \mu_j, \mu_{j-1} \rangle$ forms an inducing path that is shorter than μ. But then we reach a contradiction because

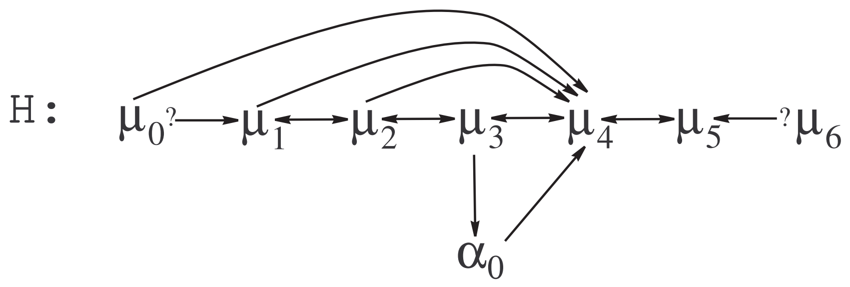


Figure 2.23: Diagram for Proof of Lemma 2.6.1 with $j = 3$. See text for details.

$\langle \mu_n, \mu_{j-1}, \ldots, \mu_0 \rangle$ is a shorter minimal inducing path between μ_0 and μ_n than μ.

A symmetric argument can be made for the case in which $\mu_j \leftarrow \alpha_0 \leftarrow \mu_{j+1}$ in $\mathcal{H}$. So, the configurations $\mu_j \rightarrow \alpha_0 \rightarrow \mu_{j+1}$ and $\mu_j \leftarrow \alpha_0 \leftarrow \mu_{j+1}$ cannot occur on a minimal inducing path in $\mathcal{H}$ for $1 < j < n - 1$. $\square$

2.6.2 Showing $\mu_i \rightarrow \alpha_0$ edge is Real

I first prove the case in which $n = 2$, and then the case for which $n > 2$.

Lemma 2.6.2 *Let $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Suppose the shortest minimal inducing path with non-adjacent endpoints in $\mathcal{H}$ is of length 2 and label the nodes $\langle \mu_0, \mu_1, \mu_2 \rangle$. Call this path μ. Let α_0 be the first descendant of μ_1 on the ancestral path from μ_1 to an endpoint in $\mathcal{H}$. Then the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real.*

Proof:

Suppose for a contradiction that the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is not real. Then there is at least one ancestral graph $\mathcal{G}$ that gave rise to $\mathcal{H}$ in which $\mu_1 \leftrightarrow \alpha_0$ occurs. Since all graphs that gave rise to $\mathcal{H}$ are Markov equivalent, by Theorem 2.1.1 the $\langle \mu_1, \alpha_0 \rangle$ edge is shielded from both sides; i.e. μ_0 and μ_2 are each adjacent to α_0: if μ_i, $i = 0, 2$ were not adjacent to α_0 then $\langle \mu_i, \mu_1, \alpha_0 \rangle$ forms an unshielded non-collider that is present in all Markov equivalent $\mathcal{G}$ that gave rise to $\mathcal{H}$, hence the $\mu_i \rightarrow \alpha_0$ edge would be real.

There is at least one ancestral $\mathcal{G}$ that gave rise to $\mathcal{H}$ such that $\mu_1 \to \alpha_0$. In this $\mathcal{G}$, where μ_1 is a parent of α_0, μ_0 and μ_2 are either parents or spouses of α_0 because otherwise either $\mu_0? \to \mu_1 \to \alpha_0 \to \mu_0$ or $\mu_2 \leftarrow \alpha_0 \leftarrow \mu_1 \leftarrow ?\mu_2$ would form a non-ancestral configuration (see Figure 2.24). In other words, whenever $\mu_1 \to \alpha_0$ occurs in some $\mathcal{G}$, $\mu_0? \to \alpha_0 \leftarrow ?\mu_2$ also occurs. Note that $\mu_0? \to \alpha_0 \leftarrow ?\mu_2$ cannot occur in $\mathcal{H}$ because if it did, then the minimality of the path μ would be violated. Since α_0 is sometimes a collider and sometimes a non-collider on the path $\langle \mu_0, \alpha_0, \mu_2 \rangle$ in the ancestral graphs that gave rise to $\mathcal{H}$, μ_0 and μ_2 are adjacent by Theorem 2.1.1 (since Markov equivalent graphs share the same unshielded colliders). But this is a contradiction. Therefore, the $\mu_1 \to \alpha_0$ edge is real. □

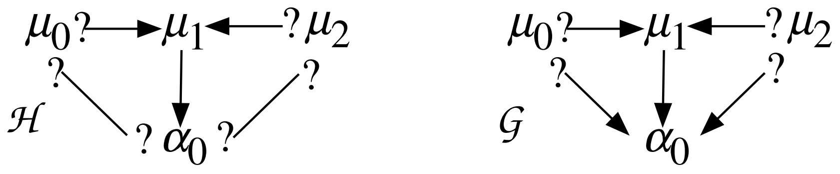


Figure 2.24: If $n = 2$ and the path μ is as outlined in Lemma 2.6.2, then μ_0 and μ_2 are adjacent to α_0.

Corollary 2.6.1 *(to Lemma 2.5.5) Let μ_i be an interior node on a minimal inducing path in some joined graph $\mathcal{H}$, and α_0 be the first descendant of μ_i on the ancestral path from μ_i to an endpoint, as shown in Figure 2.25. If α_0 occurs on the path μ, then the $\mu_i \to \alpha_0$ edge is real. Note that by Corollary 2.5.2, $\mu_0 \neq \alpha_0 \neq \mu_n$.*

Proof:

This result follows directly from Lemma 2.5.5 because $\mu_{i-1} \neq \alpha_0 \neq \mu_{i+1}$, and by Lemma 2.5.5, any edges connecting (non-consecutive) vertices on the path μ are real. □

Lemma 2.6.3 *Let $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent ancestral graphs $\mathcal{G}$. Consider the shortest minimal inducing path with non-adjacent endpoints in $\mathcal{H}$; label the nodes $\langle \mu_0, \mu_1, \ldots, \mu_{n-1}, \mu_n \rangle$, and call this path μ. Let α_0 be the first descendant of*

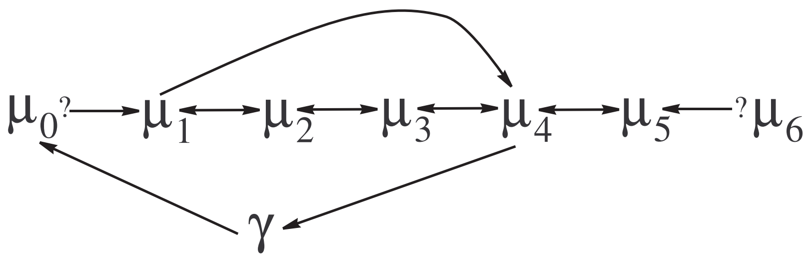


Figure 2.25: Diagram for Proof of Corollary 2.6.1. See text for details.

μ_1 on the ancestral path from μ_1 to an endpoint in $\mathcal{H}$ with $\mu_0 ? - \alpha_0$. Then the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real.

Proof:

Suppose for a contradiction that the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is not real. Then there is at least one ancestral graph, $\mathcal{G}$, that has an arrowhead at μ_1 on the $\mu_1 \rightarrow \alpha_0$ edge. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_1 \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_0 and μ_2 are each adjacent to α_0.

Case I: $n = 2$, i.e. $\mu_0 ? \rightarrow \mu_1 \leftarrow ? \mu_2$ is in $\mathcal{H}$.

By Lemma 2.6.2 the edge $\mu_1 \rightarrow \alpha_0$ in $\mathcal{H}$ is real which is a contradiction.

Case II: $n \geq 3$.

By Corollary 2.6.1, $\alpha_0 \neq \mu_t, t \neq i$. I will show that for each $\mu_r, 2 \leq r \leq (n-1)$:

1. $\alpha_0 ? - \mu_r$ in $\mathcal{H}$.
2. $\mu_0 \leftarrow \mu_r$ in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

Inductive Case: $2 \leq r < n$ Let μ_r be the first vertex after μ_1 such that the following two conditions do *not* hold:

1. $\alpha_0 ? - \mu_r$ in $\mathcal{H}$.

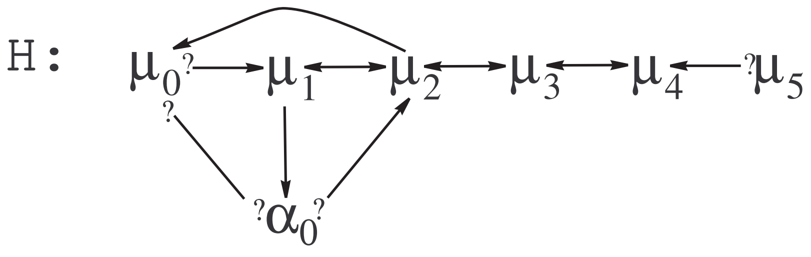


Figure 2.26: Diagram (a) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

2. $\mu_0 \leftarrow \mu_r$ in $\mathcal{H}$ and this edge is real.

If μ_r is not adjacent to α_0 and $r = 2$ then $\langle \mu_0, \mu_1, \alpha_0 \rangle$ forms an unshielded collider. If μ_r is not adjacent to α_0 and $r > 2$, then $\langle \mu_r, \mu_{r-1}, \ldots, \mu_2, \mu_1, \alpha_0 \rangle$ forms a discriminating path for μ_1 and this path is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, which is a contradiction since by Lemma 2.1.1 $\mu_1 \to \alpha_0$ would be real. Therefore μ_r is adjacent to α_0.

I will now show that $\alpha_0? - \mu_r$ is in $\mathcal{H}$. Suppose for a contradiction that $\alpha_0? \to \mu_r$ is in $\mathcal{H}$.

(a) For $r < n$ if $\alpha_0 \to \mu_r$ is in $\mathcal{H}$, then $\langle \mu_1, \mu_2, \ldots, \mu_r \rangle$ forms an inducing path in $\mathcal{H}$ that is shorter than μ (because each μ_i, $2 \leq i < r$ is a parent of μ_0, and $\mu_1 \to \alpha_0 \to \mu_r$ is in $\mathcal{H}$) which would imply that μ_r is adjacent to μ_0, and by Corollary 2.5.2 $\mu_0 \leftarrow \mu_r$ is in $\mathcal{H}$ and this edge is real. See Figure 2.27.

Consider the case where $r = n$. If $\alpha_0 \to \mu_r$ is in $\mathcal{H}$ then since μ_0 and μ_r are not adjacent, $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider that is present in all $\mathcal{G}$ joined to form $\mathcal{H}$. By the definition of μ_r each of $\{\mu_2, \ldots, \mu_{n-1}\}$ is a parent of μ_0 in every $\mathcal{G}$ joined to form $\mathcal{H}$. Consider the maximal ancestral graph $\mathcal{G}^*$, used to form $\mathcal{H}$, in which $\mu_1 \to \alpha_0$ (See Figure 2.28); such a $\mathcal{G}$ exists because $\mu_1 \to \alpha_0$ in $\mathcal{H}$. In $\mathcal{G}^*$, $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider so α_0 is a parent of at least one of μ_0 or μ_n. In fact, by Lemma 2.5.3 $\mu_0? \to \alpha_0$ occurs in $\mathcal{G}^*$ so $\alpha_0 \to \mu_n$ occurs in $\mathcal{G}^*$. But then μ forms an inducing path in $\mathcal{G}^*$ and by the maximality of $\mathcal{G}^*$, μ_0 is adjacent to μ_n. Consequently μ_0 is adjacent to μ_n in $\mathcal{H}$, which is a contradiction. Therefore if $\alpha_0 \to \mu_r$ is in $\mathcal{H}$, then $r < n$.

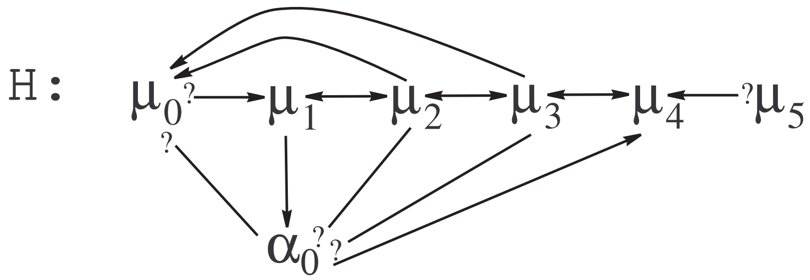


Figure 2.27: Diagram (b) for Proof of Lemma 2.6.3 with $r = n = 5$. See text for details.

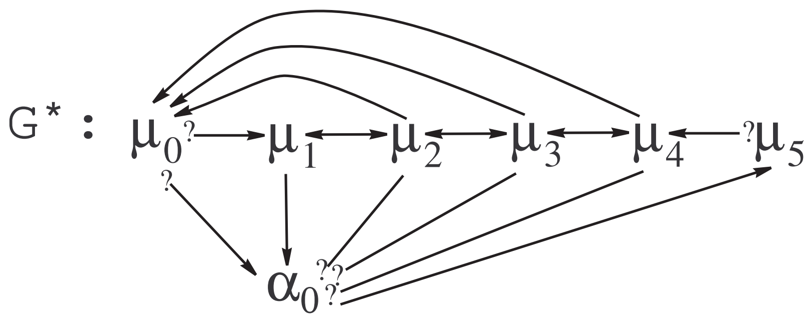


Figure 2.28: Diagram (c) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

(b) If $\alpha_0 \leftrightarrow \mu_r$ is in $\mathcal{H}$, then $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a non-collider in $\mathcal{H}$ (with an arrowhead at α_0 on the $\{\alpha_0, \mu_r\}$ edge in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$). We know that there is at least one $\mathcal{G}$ in which $\mu_0? \rightarrow \alpha_0$ occurs and in this graph $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a collider. We also know that there is at least one $\mathcal{G}^**$ in which $\mu_0 \leftarrow \alpha_0$ and hence $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a non-collider in $\mathcal{G}^**$. Consequently, So μ_0 is adjacent to μ_r (see Figure 2.29). If $r = n$ then we reach a contradiction because by assumption μ_0 and μ_n are not adjacent. Thus, if $\alpha_0 \leftrightarrow \mu_r$ then $r < n$ and by Corollary 2.5.2, $\mu_0 \leftarrow \mu_r$ and this edge is real.

So, if $\alpha_0? \rightarrow \mu_r$ is in $\mathcal{H}$, then $r < n$, $\mu_0 \leftarrow \mu_r$ and this edge is real. But then $\langle \mu_{r+1}, \mu_r, \alpha_0, \mu_0 \rangle$ forms a discriminating path for α_0 in $\mathcal{H}$, and discriminates $\langle \mu_0, \alpha_0, \mu_r \rangle$

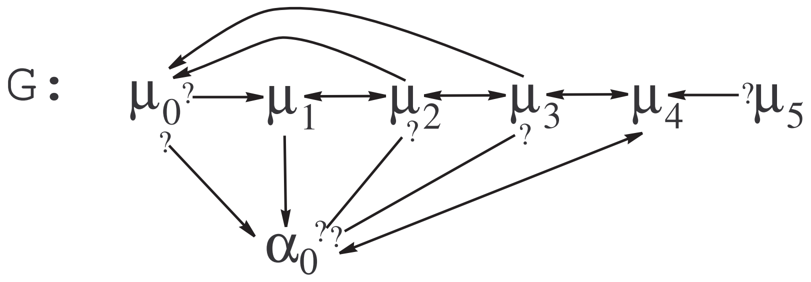


Figure 2.29: Diagram (d) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. However, this is a contradiction for the following reason:

Consider a $\mathcal{G}^*$ in which $\mu_1 \rightarrow \alpha_0$ (see Figure 2.30). By Lemma 2.5.3 (triangle rule), $\mu_0? \rightarrow \alpha_0$ is in $\mathcal{G}^*$. Because the $\mu_0 \leftarrow \mu_r$ edge is real, by Lemma 2.5.3 there is also an arrowhead at α_0 on the $\{\alpha_0, \mu_r\}$ edge, i.e. $\alpha_0 \leftrightarrow \mu_r$. But then $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a collider in $\mathcal{G}^*$, which is a contradiction.

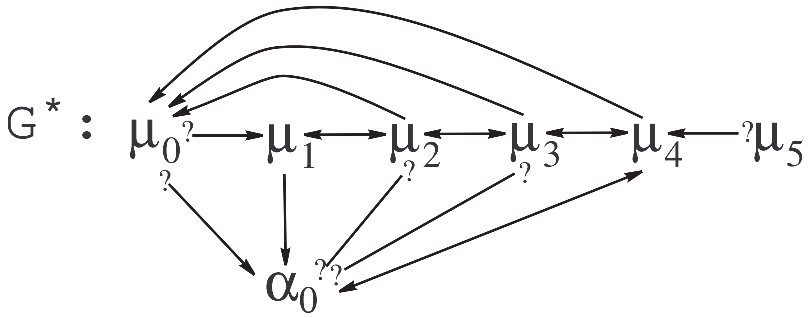


Figure 2.30: Diagram (e) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

Therefore, I have shown that $\alpha_0? \rightarrow \mu_r$ is not in $\mathcal{H}$. Consequently we can infer that $\alpha_0? - \mu_r$ is in $\mathcal{H}$. I will now show that μ_r is adjacent to μ_0 (which leads to an immediate contradiction because of the definition of μ_r).

Currently we know the following about $\mathcal{H}$ (see Figure 2.31):

a) $\mu_0 ? \!-\! \alpha_0$,

b) $\mu_0 \leftarrow \mu_i$ for $2 \leq i < r$ and these edges are real, and

c) $\alpha_0 ? \!-\! \mu_i$ for $2 \leq i \leq r$

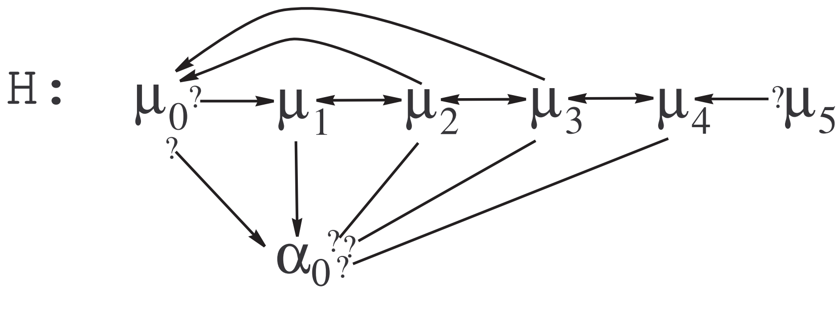


Figure 2.31: Diagram (f) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

Suppose for a contradiction that μ_r is not adjacent to μ_0. Then $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider in $\mathcal{H}$ and this non-collider is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

Consider the $\mathcal{G}$ in which $\mu_1 \rightarrow \alpha_0$ occurs. In this $\mathcal{G}$, $\mu_0 ? \!\rightarrow \alpha_0$ also occurs, so $\mu_i \rightarrow \mu_0 ? \!\rightarrow \alpha_0 ? \!-\! ?\mu_i$ occurs in $\mathcal{H}$ for $2 \leq i < r$. By a simple triangle argument, and the fact that $\mathcal{G}$ is ancestral, we can infer an arrowhead at α_0 on the $\{\mu_i, \alpha_0\}$ edge for every i, $2 \leq i < r$. I have already shown that $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a non-collider in all graphs that gave rise to $\mathcal{H}$. Since $\mu_0 ? \!\rightarrow \alpha_0$ is in $\mathcal{G}$ we can infer that $\alpha_0 \rightarrow \mu_r$ is in $\mathcal{G}$ (else $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a collider). See Figure 2.32.

If there is no arrowhead at μ_1 on the $\{\mu_1, \alpha_0\}$ edge in $\mathcal{G}$, i.e. $\mu_1 \rightarrow \alpha_0$, then $\pi(\mu_0, \mu_r)$ forms an inducing path and by the maximality of $\mathcal{G}$, μ_r would be adjacent to μ_0 which would be a contradiction. Therefore $\mu_1 \leftrightarrow \alpha_0$ occurs in this $\mathcal{G}$.

So, in any $\mathcal{G}$ in which $\mu_0 ? \!\rightarrow \alpha_0$ occurs the following also occurs:

(i) $\alpha_0 \leftarrow ?\mu_i$ for $2 \leq i < r$,

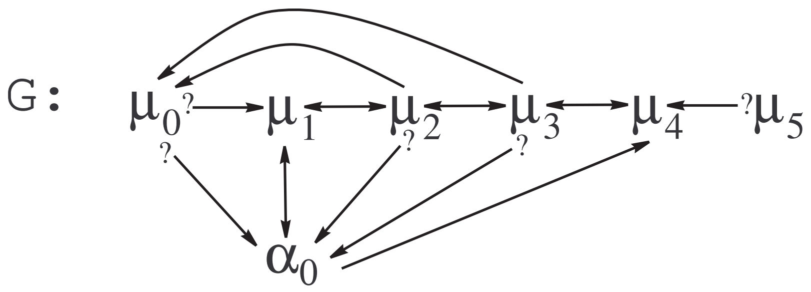


Figure 2.32: Diagram (g) for Proof of Lemma 2.6.3 with $r = 4$. See text for details.

(ii) $\alpha_0 \rightarrow \mu_r$, and

(iii) $\mu_1 \leftrightarrow \alpha_0$.

But this is a contradiction because: (a) it implies that whenever $\langle \mu_0, \alpha_0, \mu_2 \rangle$ forms a collider in $\mathcal{G}$, $\mu_1 \leftrightarrow \alpha_0$ also occurs and (b) I have already argued that whenever $\langle \mu_0, \alpha_0, \mu_2 \rangle$ forms a non-collider then $\mu_1 \leftrightarrow \alpha_0$ also occurs. In other words, if μ_0 is not adjacent to μ_r then $\mu_1 \leftrightarrow \alpha_0$ occurs in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$, and hence $\mu_1 \leftrightarrow \alpha_0$ is in H which is a contradiction. But then μ_0 is adjacent to μ_r. If $r = n$ then we reach an immediate contradiction because by assumption μ_0 and μ_n are not adjacent. If $r < n$ then by Corollary 2.5.2 $\mu_0 \leftarrow \mu_r$ is in $\mathcal{H}$ and this edge is real. But this too is a contradiction because of the definition of μ_r.

Therefore I have shown that if the $\mu_1 \rightarrow \alpha_0$ edge is not real, then we reach some contradiction. Therefore, the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real. $\square$

Lemma 2.6.4 *Let $\pi = \langle \mu_0, \ldots, \mu_2 \rangle$ be the shortest minimal inducing path with non-adjacent endpoints in the graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Let α_0 be the first descendant of μ_1 on the ancestral path from μ_1 to Z in $\mathcal{H}$ with $\mu_0? \rightarrow \alpha_0$. Then the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real.*

Proof:

Suppose for a contradiction that the $\mu_1 \rightarrow \alpha_0$ edge is not real. Then there is at least one ancestral graph $\mathcal{G}$ in which the $\{\mu_1, \alpha_0\}$ edge is oriented $\mu_1 \leftrightarrow \alpha_0$. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_1 \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_0 and μ_2 are each adjacent to α_0.

Case I: $n = 2$, i.e. $\mu_0? \rightarrow \mu_1 \leftarrow ?\mu_2$ is in $\mathcal{H}$. By Lemma 2.6.2 the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real which is a contradiction.

Case II: $n \geq 3$. By Corollary 2.6.1, α_0 is not equal to another $\mu_t, t \neq i$. Since μ is minimal, $\alpha_0 \leftrightarrow \mu_j$, $2 \leq j \leq n$ does not occur in $\mathcal{H}$ because in this case $\langle \mu_0, \alpha_0, \mu_j, \ldots, \mu_n \rangle$ would violate the minimality of μ. Let μ_k be the vertex furthest from μ_1 such that μ_i is adjacent to α_0 for $0 \leq i \leq k$. Since μ_2 is adjacent to α_0, $r \geq 2$.

Claim: $\alpha_0 \leftarrow \mu_i$ is in $\mathcal{H}$ for $1 \leq i \leq k$.
Suppose that there exists an $i \leq k$ such that $\alpha_0 \,—\, ?\mu_k$ is in $\mathcal{H}$. Let μ_r be the vertex closest to μ_1 such that $\alpha_0 \,—\, ?\mu_r$ is in $\mathcal{H}$. By the definition of μ_r, $\alpha_0 \leftarrow \mu_{r-1}$ is in $\mathcal{H}$. If $r = 2$, then by assumption $\alpha_0 \leftarrow \mu_1$ and the claim is true. There are now two cases to consider:

(i) μ_r is adjacent to μ_0 (See Figure 2.33).

If μ_r is adjacent to μ_0, then by Corollary 2.5.2 $\mu_0 \leftarrow \mu_r$, and this edge is real. Hence by Lemma 2.5.3, since $\mu_0? \rightarrow \alpha_0 \,—\, ?\mu_r \rightarrow \mu_0$ it follows that there is an arrowhead at α_0 on the $\{\mu_r, \alpha_0\}$ edge in $\mathcal{H}$ (i.e. $\alpha_0 \leftarrow ?\mu_r$), which is a contradiction. Therefore, μ_r is not adjacent to μ_0.

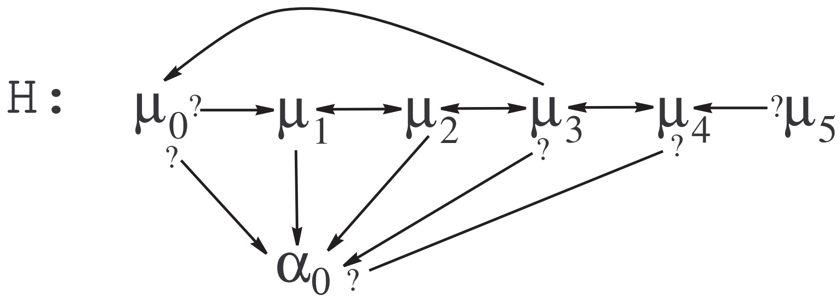


Figure 2.33: Diagram (a) for Proof of Lemma 2.6.4 with $r = 4$. See text for details.

(ii) μ_r is not adjacent to μ_0 (See Figure 2.34).

In this case, $\mu_0 ? \rightarrow \alpha_0 - ?\mu_r$ forms an unshielded non-collider that is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Hence $\alpha_0 \rightarrow \mu_r$ is in $\mathcal{H}$ and because by definition of the join operation $\mu_0 ? \rightarrow \alpha_0$ occurs in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$, the $\alpha_0 \rightarrow \mu_r$ edge in $\mathcal{H}$ is real. Since $\alpha_0 \leftarrow \mu_{r-1}$ is in $\mathcal{H}$ (by the definition of μ_r), there exists at least one maximal ancestral graph $\mathcal{G}^*$ in which $\alpha_0 \leftarrow \mu_{r-1}$. In this graph $\mathcal{G}^*$, the triangle $\mu_r ? \rightarrow \mu_{r-1} \rightarrow \alpha_0 \rightarrow \mu_r$ occurs, which is a contradiction because $\mathcal{G}^*$ is ancestral.

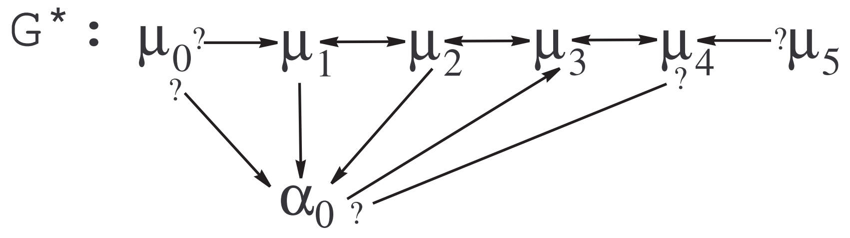


Figure 2.34: Diagram (b) for Proof of Lemma 2.6.4 with $r = 4$. See text for details.

Consequently for $1 \leq i \leq k$, $\mu_k \rightarrow \alpha_0$ occurs in $\mathcal{H}$. If $k < n$, then μ_{k+1} exists and is not adjacent to α_0. However, $\langle \mu_0, \mu_{k+1}, \ldots, \mu_2, \mu_1, \alpha_0 \rangle$ forms a discriminating path for μ_1 in $\mathcal{H}$ and by Lemma 2.5.4, this path is discriminating in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$. Hence $k = n$, which is a contradiction because $\mu_0 ? \rightarrow \alpha_0 \leftarrow ?\mu_n$ violates the minimality of the path μ. Consequently, the $\mu_1 \rightarrow \alpha_0$ edge is real.

□

Lemma 2.6.5 *Let $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent ancestral graphs $\mathcal{G}$. Consider the shortest minimal inducing path in $\mathcal{H}$; label the nodes $\langle \mu_0, \mu_1, \ldots, \mu_{n-1}, \mu_n \rangle$, and call this path μ. Let α_0 be the first descendant of μ_i on the shortest ancestral path from μ_i to an endpoint, $0 < i < n$. Then α_0 is a child of μ_i in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ (i.e. $\mu_i \rightarrow \alpha_0$ is real).*

Proof:

For a contradiction, suppose $\mu_i \rightarrow \alpha_0$ is not real. Then there is at least one ancestral graph

$\mathcal{G}$ that has an arrowhead at μ_i on the $\mu_i \rightarrow \alpha_0$ edge. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_i \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_{i-1} and μ_{i+1} are also adjacent to α_0.

Case I: $n = 2$. By Lemma 2.6.2 if $n = 2$, then the $\mu_1 \rightarrow \alpha_0$ edge is real.

Case II: $n \geq 3$. For $n \geq 3$, Lemmas 2.6.3 and 2.6.4 show that if $i = 1, (n-1)$, then the $\mu_i \rightarrow \alpha_0$ edge is real. Now I consider the case in which $n \geq 3, 1 > i > (n-1)$. Note that by Corollary 2.6.1 $\alpha_0 \neq \mu_t, t \neq i$. By Lemma 2.6.1, neither μ_{i-1} nor μ_{i+1} are children of α_0. I will show (see Figure 2.35):

1. Either there exists a $j < i$ such that $\mu_{j-1} \leftarrow ?\alpha_0, \mu_r - ?\alpha_0$, for all $r, 0 < j \leq r \leq i$ in $\mathcal{H}$, or $\mu_0 ? \rightarrow \alpha_0, \mu_r - ?\alpha_0, 0 < r < i$ in $\mathcal{H}$.

2. Either there exists a $k > i$ such that $\alpha_0 ? \rightarrow \mu_{k+1}, \alpha_0 ? - \mu_s, i < s < k < n$ in $\mathcal{H}$, or $\alpha_0 \leftarrow ?\mu_n, \mu_s - ?\alpha_0$, for all $s, i \leq s \leq n$ in $\mathcal{H}$.

3. Finally, I will show that the configurations described in 1. and 2. lead to some contradiction.

1. If $\mu_{i-1} \leftarrow ?\alpha_0$ is in $\mathcal{H}$, then $j = i$ and we are done. So, $\mu_{i-1} - ?\alpha_0$ is in $\mathcal{H}$. If μ_{i-2} is not adjacent to α_0, then $\langle \mu_{i-2}, \mu_{i-1}, \alpha_0 \rangle$ is an unshielded non-collider that is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Consequently, $\langle \mu_{i-2}, \mu_{i-1}, \mu_i, \alpha_0 \rangle$ forms a discriminating path for μ_i in $\mathcal{H}$ that is also present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, which is a contradiction. So, μ_{i-2} is adjacent to α_0.

   Let j be the smallest r in a sequence such that $\mu_r - ?\alpha_0$ in $\mathcal{H}$, for all $r, 0 < j \leq r \leq i$. If μ_{j-1} is not adjacent to α_0, then $\langle \mu_{j-1}, \mu_j, \alpha_0 \rangle$ is a non-collider in $\mathcal{H}$ and by Lemma 2.5.4 $\langle \mu_{j-1}, \mu_j, \mu_{j+1}, \alpha_0 \rangle$ discriminates μ_{j+1} to be a non-collider present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Consequently, by induction, $\langle \mu_{j-1}, \mu_j, \ldots, \mu_i, \alpha_0 \rangle$ discriminates $\langle \mu_{i-1}, \mu_i, \alpha_0 \rangle$ to be a non-collider, which is a contradiction. So, μ_{j-1} is adjacent to α_0.

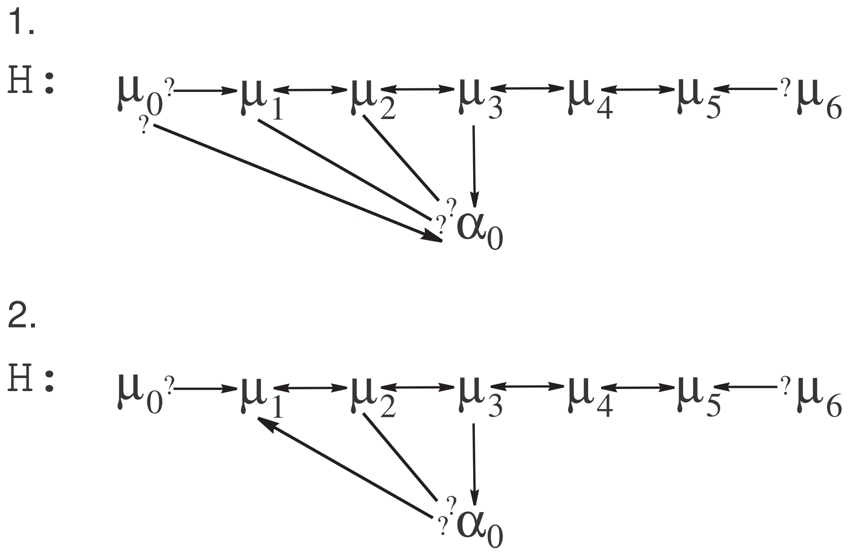


Figure 2.35: Diagram for Proof of Lemma 2.6.5. See text for details.

Consider the case in which $j = 1$. If μ_0 is not adjacent to α_0 then by induction $\langle \mu_0, \mu_1, \ldots, \mu_i, \alpha_0 \rangle$ would form a discriminating path for μ_i in $\mathcal{H}$. By Lemma 2.5.4 this path would be present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ which is a contradiction. Therefore μ_0 is adjacent to α_0. Note that the edge type at α_0 on the $\mu_0 ? \!-\! ?\alpha_0$ edge in $\mathcal{H}$ must be present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. If this were not the case, and $\langle \mu_0, \alpha_0, \mu_i \rangle$ was a collider in some $\mathcal{G}$ that gave rise to $\mathcal{H}$, but a non-collider in others, then μ_0 would be adjacent to μ_i. By Lemma 2.5.2 the $\{\mu_0, \mu_1\}$ edge is oriented $\mu_0 \leftarrow \mu_i$ in $\mathcal{H}$, which is a contradiction because the minimality of the ancestral path from μ_i to an endpoint would be violated. So, if $j = 1$, then either $\mu_0 \leftarrow \alpha_0$, or $\mu_0 ? \rightarrow \alpha_0$ in $\mathcal{H}$. If $j > 1$, then $\mu_{j-1} \leftarrow ?\alpha_0$ is in $\mathcal{H}$.

So μ_i is the end of a sequence of vertices $< \mu_j, \mu_{j+1}, \ldots, \mu_i >$ such that $\mu_r - ?\alpha_0, j \leq r \leq i$. Either $\mu_{j-1} \leftarrow ?\alpha_0 ? - \mu_j$, $0 \leq j < i$, or $\mu_0 \rightarrow \alpha_0 ? - \mu_1$ (in $\mathcal{H}$), as required.

2. A symmetric argument can be made for this case:

i.e. μ_i is the beginning of a sequence of vertices $< \mu_i, \ldots, \mu_{k-1}, \mu_k >$ such that

$\alpha_0 ? - \mu_s, i \leq s \leq k$. This sequence ends with either $\mu_{k-1} - ?\alpha_0 ? \rightarrow \mu_k$, $i < k \leq n$ or $\mu_{n-1} - ?\alpha_0 \leftarrow \mu_n$ (in $\mathcal{H}$).

Note that if $\mu_0 ? \rightarrow \alpha_0$ is in $\mathcal{H}$, then $\alpha_0 \leftarrow ?\mu_n$ is not in $\mathcal{H}$ (and vice versa) because otherwise $\mu_0 ? \rightarrow \alpha_0 \leftarrow ?\mu_n$ would form an inducing path with non-adjacent endpoints that is shorter than the original path μ, contrary to the minimality of the path. Similarly, if $\mu_l \leftrightarrow \alpha_0, 0 < l < i$ is in $\mathcal{H}$, then $\alpha_0 \leftrightarrow \mu_m, i < m < n$ does not occur in $\mathcal{H}$. Without loss of generality, assume that $\mu_l \leftrightarrow \alpha_0, 0 < l < i$ does not occur in $\mathcal{H}$ (the other case can be argued in a symmetrical fashion). Then there are 3 cases left to consider (see Figure 2.36):

a) $\mu_{j-1} \leftarrow \alpha_0, (0 < j < i)$, $\mu_t - ?\alpha_0$, for all $t, (j \leq t < n)$, $\mu_i \rightarrow \alpha_0$, and $\alpha_0 \leftarrow ?\mu_n$ in $\mathcal{H}$,

b) $\mu_0 \rightarrow \alpha_0$, $\mu_t - ?\alpha_0, (0 < t \leq k)$, $\mu_i \rightarrow \alpha_0$ and $\alpha_0 ? \rightarrow \mu_{k+1}, (i < k < n)$ in $\mathcal{H}$, or

c) $\mu_{j-1} \leftarrow \alpha_0, (0 < j < i)$, $\mu_t - ?\alpha_0$, for all $t, (j \leq t \leq k)$, $\mu_i \rightarrow \alpha_0$, and $\alpha_0 ? \rightarrow \mu_{k+1}, (i < k < n)$ in $\mathcal{H}$.

I will show that each of these configurations leads to some contradiction.

3. a) $\mu_{j-1} \leftarrow \alpha_0, (0 < j < i)$, $\mu_t - ?\alpha_0$, for all $t, (j \leq t < n)$, $\mu_i \rightarrow \alpha_0$, and $\alpha_0 \leftarrow ?\mu_n$ in $\mathcal{H}$.

The tail at α_0 on the $\mu_{j-1} \leftarrow \alpha_0$ edge is not present in all ancestral $\mathcal{G}$ that gave rise to $\mathcal{H}$, else $\mu_j \leftarrow ?\alpha_0$ would also occur in $\mathcal{H}$ which is a contradiction. So, μ_{j-1} is adjacent to μ_n, (and possibly more vertices). By Corollary 2.5.2, $\mu_{j-1} \rightarrow \mu_n$ is in $\mathcal{H}$ and this edge is real. Since $\mu_n ? \rightarrow \alpha_0 \rightarrow \mu_{j-1} \rightarrow \mu_n$ is in $\mathcal{H}$ and the $\mu_{j-1} \rightarrow \mu_n$ edge being real, by Lemma 2.5.2, $\mu_n \leftrightarrow \alpha_0 \leftrightarrow \mu_{j-1} \rightarrow \mu_n$ would be a subgraph of $\mathcal{H}$, which is a contradiction (because $\langle \mu_0, \ldots, \mu_{j-1}, \alpha_0, \mu_n \rangle$ would form a shorter minimal inducing path with non-adjacent endpoints that is shorter than μ).

b) $\mu_0 \rightarrow \alpha_0$, $\mu_t - ?\alpha_0$, for all $t, (0 < t \leq k)$, $\mu_i \rightarrow \alpha_0$ and $\alpha_0 ? \rightarrow \mu_{k+1}, (i < k < n)$ in $\mathcal{H}$.

If $\alpha_0 \leftrightarrow \mu_{k+1}$ is in $\mathcal{H}$, then $\langle \mu_0, \alpha_0, \mu_{k+1}, \ldots, \mu_n \rangle$ is an inducing path with non-adjacent

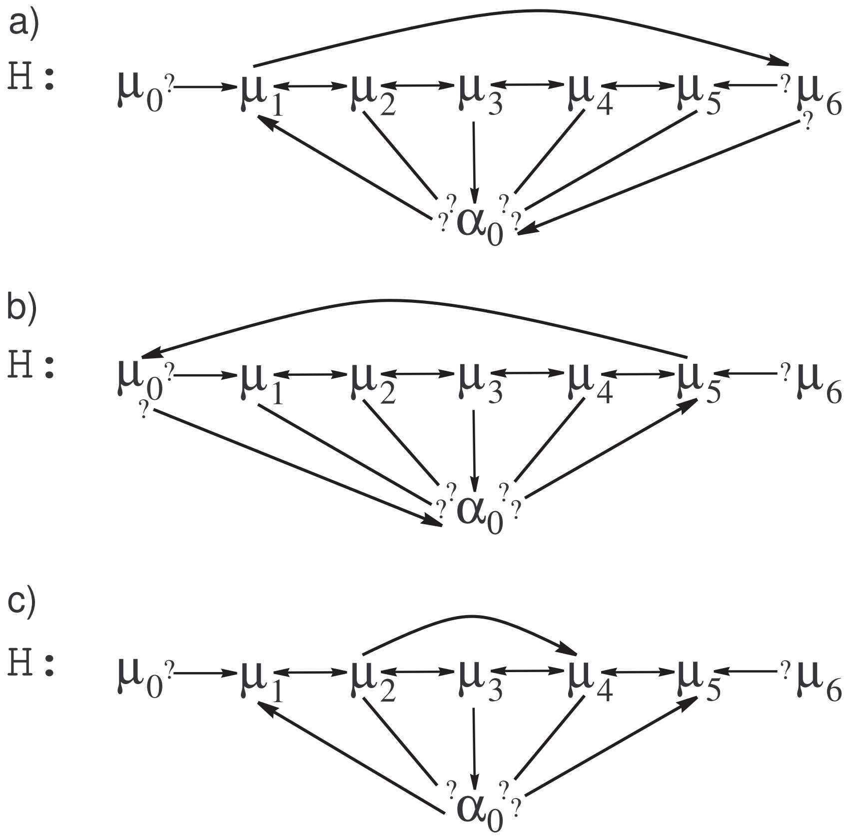


Figure 2.36: Diagram for Proof of Lemma 2.6.5. See text for details.

endpoints that is more minimal than μ. If $\alpha_0 \to \mu_{k+1}$ is in $\mathcal{H}$, then an argument analogous to that given in a) leads to a contradiction. In particular, the tail at α_0 on the $\alpha_0 \to \mu_{k+1}$ edge is not real so μ_{k+1} is adjacent to μ_0 (and possibly more vertices). By Corollary 2.5.2, $\mu_0 \leftarrow \mu_{k+1}$ is in $\mathcal{H}$ and this edge is real. Since $\mu_0 ? \to \alpha_0 \to \mu_{k+1} \to \mu_0$ is in $\mathcal{H}$ with the $\mu_{k+1} \leftarrow \mu_0$ edge is real, by Lemma 2.5.2 $\mu_0 \leftrightarrow \alpha_0 \leftrightarrow \mu_{k+1} \to \mu_0$ would be a subgraph of $\mathcal{H}$, which is a contradiction (because $\langle \mu_0, \alpha_0, \mu_{k+1}, \ldots, \mu_n \rangle$ would form a minimal inducing path with non-adjacent endpoints that is shorter than μ).

c) $\mu_{j-1} \leftarrow \alpha_0, (0 < j < i)$, $\mu_t - ?\alpha_0$, for all $t, (j \leq t \leq k)$, $\mu_i \to \alpha_0$, and $\alpha_0 ? \to \mu_{k+1}, (i < k < n)$ in $\mathcal{H}$.

Since $\mu_i \to \alpha_0$ is in $\mathcal{H}$, there exists at least one ancestral graph, $\mathcal{G}$, that gave rise to $\mathcal{H}$ for which $\mu_i \to \alpha_0$. In this ancestral graph, $\langle \mu_{i-1}, \alpha_0, \mu_{i+1} \rangle$ forms a collider since otherwise $\mathcal{G}$ would not be ancestral, i.e. $\mu_{i-1} \leftrightarrow \mu_i \to \alpha_0 \to \mu_{i-1}$ or $\mu_{i+1} \leftrightarrow \mu_i \to \alpha_0 \to \mu_{i+1}$ would occur in $\mathcal{G}$. Because $\langle \mu_{i-1}, \alpha_0, \mu_{i+1} \rangle$ is a collider in at least one $\mathcal{G}$ that gave rise to $\mathcal{H}$, but is a non-collider in $\mathcal{H}$, μ_{i-1} and μ_{i+1} are adjacent.

There are now two sub-cases left to consider:

i) If $\mu_{i-1} \to \mu_{i+1}$ in $\mathcal{H}$ then by Lemma 2.5.5 this edge is real. Now consider the ancestral graph $\mathcal{G}^*$ in which $\alpha_0 \leftarrow \mu_{i+1}$ (such a graph exists since $\mu_{i+1} -?\alpha_0$ is in $\mathcal{H}$). Note that $\mu_{i-1} \to \alpha_0$ is in $\mathcal{G}^*$ because $\alpha_0 ? \to \mu_{i-1} \to \mu_{i+1} \to \alpha_0$ is not ancestral. Consequently, by Lemma 2.6.1, $\mu_{i-2} ? \to \alpha_0$ and $\langle \mu_{i-2}, \alpha_0, \mu_{i+1} \rangle$ forms a collider not present in $\mathcal{H}$. So, just as in the case of μ_{i-1}, μ_{i-2} is adjacent to μ_{i+1}. By Corollary 2.5.3 $\mu_{i-2} \to \mu_{i+1}$ is in $\mathcal{H}$ and this edge is real. Because $\alpha_0 ? \to \mu_{i-2} \to \mu_{i+1} \to \alpha_0$ is not ancestral $\mu_{i-2} \to \alpha_0$ is in $\mathcal{G}^*$. By induction, all $\mu_r, 0 \leq j-1 \leq r < i$ are adjacent to and parents of μ_{i+1} and α_0. But we now reach a contradiction because $\alpha_0 ? \to \mu_{j-1} \to \mu_{i+1} \to \alpha_0$ is not ancestral.

ii) If $\mu_{i+1} \to \mu_{i-1}$ in $\mathcal{H}$ then by Lemma 2.5.5 this edge is real. Now consider the ancestral graph $\mathcal{G}^{**}$ in which $\mu_{i-1} \to \alpha_0$. In this ancestral graph, $\mu_{i+1} \to \alpha_0$ because $\alpha_0 ? \to \mu_{i+1} \to \mu_{i-1} \to \alpha_0$ is not ancestral. Consequently, $\mu_{i+2} ? \to \alpha_0$ and $\langle \mu_{i+2}, \alpha_0, \mu_{i-1} \rangle$ forms a collider not present in $\mathcal{H}$. So, just as in the case of μ_{i+1}, μ_{i+2} is adjacent to μ_{i-1}. And because $\alpha_0 ? \to \mu_{i+2} \to \mu_{i-1} \to \alpha_0$ is not ancestral $\mu_{i+2} \to \alpha_0$ in $\mathcal{G}^{**}$. By induction, all $\mu_s, i < s \leq k+1 \leq n$ are adjacent to and parents of μ_{i-1} and α_0. But we now reach a contradiction because $\alpha_0 ? \to \mu_{k+1} \to \mu_{i-1} \to \alpha_0$ is not ancestral.

Since any joined graph $\mathcal{H}$, for which the edge $\mu_i \to \alpha_0$ is *not* real leads to a contradiction, the edge $\mu_i \to \alpha_0$ is real. $\square$

2.6.3 Main Result 1: Maximality of Joined Graphs

I now prove the main result of this section:

Theorem 2.6.1 *(Main Result 1: Maximality of Joined Graphs) Let $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent ancestral graphs $\mathcal{G}$: If all $\mathcal{G}$ are maximal, then $\mathcal{H}$ is also maximal.*

Proof:

For a contradiction, let us assume that $\mathcal{H}$ is not maximal. Then, there must be at least one pair of non-adjacent vertices such that there is an inducing path between them. Consider the shortest minimal inducing path with non-adjacent endpoints in $\mathcal{H}$; label the nodes $\langle \mu_0, \mu_1, \ldots, \mu_{n-1}, \mu_n \rangle$, and call this path μ.

From the definition of an inducing path, we know that μ is a collider path such that $\langle \mu_1, \ldots, \mu_{n-1} \rangle$ are each ancestors of either μ_0 or μ_n in $\mathcal{H}$. By the minimality of the path μ, we also know that there is no inducing path on a subset of μ for which the endpoints are not adjacent. Finally note that the collider path between μ_0 and μ_n is also present in all ancestral graphs, $\mathcal{G}$, that gave rise to $\mathcal{H}$.

I will show that all the edges on the directed path from μ_i to an endpoint are real. Hence, I will have shown that the inducing path in $\mathcal{H}$ is also present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are maximal, if μ is an inducing path in all $\mathcal{G}$, then μ_0 and μ_n are adjacent in all $\mathcal{G}$; hence they are adjacent in $\mathcal{H}$, and we reach a contradiction.

Let $\mu_i \to \alpha_0 \to \ldots \to \alpha_m$ be the shortest directed path from μ_i to an endpoint in $\mathcal{H}$ (so $\alpha_m = \mu_0$ or $\alpha_m = \mu_n$). It is sufficient to show that this path is present in any $\mathcal{G}$ joined to form $\mathcal{H}$ (because in that case there would be at least one ancestral graph $\mathcal{G}$ that gave rise to $\mathcal{H}$ that was not maximal, which would be a contradiction).

Lemma 2.6.5 shows that the first descendant, α_0, of any vertex along μ, say μ_i, is a child of μ_i in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ (i.e. $\mu_i \to \alpha_0$ is real). I will use an inductive proof to show that all the subsequent edges on the directed path from μ_i to an endpoint are also real.

Suppose that $\alpha_0 \to \alpha_1$ is not real (see Figure 2.37). Then by Theorem 2.1.1, $\mu_i \to \alpha_0 \to \alpha_1$ is shielded (i.e. μ_i is adjacent to α_1) since otherwise $\langle \mu_i, \alpha_0, \alpha_1 \rangle$ is an unshielded collider. By Lemma 2.5.3 $\mu_i \to \alpha_1$ is in $\mathcal{H}$ because the $\mu_i \to \alpha_0$ edge is real; but then the minimality of the path is violated since we can take a directed path from μ_i to an endpoint that bypasses α_0. So, the edge $\alpha_0 \to \alpha_1$ is real.

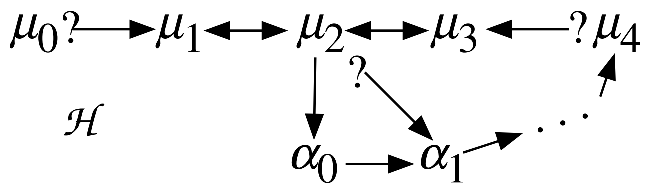


Figure 2.37: If the path μ is as outlined in Lemma 2.6.2, then μ_2 and α_1 are adjacent, and μ_2 is a parent of α_1.

Assume that for $0 < k < m$, the $\alpha_{k-1} \to \alpha_k$ edge is real and consider the $\alpha_k \to \alpha_{k+1}$ edge. If $\alpha_k \to \alpha_{k+1}$ in $\mathcal{H}$ is not real, then by Theorem 2.1.1, α_{k-1} is adjacent to α_{k+1}. By Lemma 2.5.3, $\alpha_{k-1} \to \alpha_{k+1}$ occurs in $\mathcal{H}$ (because the $\alpha_{k-1} \to \alpha_k$ edge is real), but then the minimality of the path is violated since we can take the path from μ_i to an endpoint that bypasses α_k. Consequently, the $\alpha_k \to \alpha_{k+1}$ edge is real. So, by induction, the edges between $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m$ are real and form a directed path from α_0 to an endpoint.

I have now shown that μ also occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Since all these ancestral graphs $\mathcal{G}$ are maximal, μ_0 and μ_n are adjacent in every $\mathcal{G}$. Consequently, μ_0 and μ_n are adjacent on $\mathcal{H}$ which is a contradiction. □

In general, there is no pairwise Markov property for joined graphs. Theorem 2.6.1 tells us that if we restrict ourselves to the joining sets of Markov equivalent ancestral graphs for which the pairwise Markov property holds, i.e. maximal ancestral graphs, then the resulting graph is also maximal. However, to prove that such joined graphs follow a pairwise Markov property, it is necessary to show that for paths with non-adjacent endpoints there exists a set that j-separates the endpoints given Z. Chapter 3 will prove that a global Markov property holds for joined graphs formed by joining sets of Markov equivalent ancestral graphs. The pairwise Markov property follows as a corollary to this result (see Corollary 3.8.1).

Chapter 3

GLOBAL MARKOV PROPERTY FOR JOINED GRAPHS

In this chapter I investigate the general structure of minimal j-connecting paths in joined graphs and prove the Global Markov property for joined graphs. In Lemma 3.2.1, I show that the edges connecting non-consecutive vertices along a minimal j-connecting path given Z (see Definition 3.1.1) in a joined graph are directed out of a collider in Z and this edge is real. We will then use this fact to show that if there is a minimal j-connecting path between α and β in a joined graph given some separating set Z, then there exists an m-connecting path given Z between α and β in all ancestral graphs that gave rise to $\mathcal{H}$. The second main result of this thesis is presented in Theorem 3.8.1 where I show that graphs formed by joining Markov equivalent maximal ancestral graphs obey a global Markov property with respect to j-connection. An outline of this proof is given in Section 3.3 (see Figure 3.15).

3.1 Adjacencies on j-connecting Paths

Some useful facts related to the structure of j-connecting paths would be to know which non-consecutive vertices along the path are adjacent; and furthermore, if we can infer the orientation of such connecting edges. The following Lemmas provide insights to such issues for *minimal j-connecting paths* given Z, where a minimal j-connecting path given Z is defined as follows:

Definition 3.1.1 *Let μ be a j-connecting path given Z in a joined graph with vertices $< \mu_0, \mu_2 \ldots, \mu_n >$, and let ψ_i be the number of edges on a shortest directed path between μ_i and a member in Z. Furthermore, let $\phi(\mu)$ be the total number of edges between each interior vertex and the closest member of Z on these ancestral paths, i.e. $\phi(\mu) = \Sigma_{i=1}^{n-1}\psi_i$. Then, μ is a "minimal j-connecting path given Z" between vertices μ_0 and μ_n in a joined graph if:*

(i) no subsequence of the vertices on μ form a j-connecting path given Z between μ_0 and μ_n, and

(ii) there is no other path μ' which is j-connecting given Z between μ_0 and μ_n with the same number of vertices as μ such that $\phi(\mu') < \phi(\mu)$.

Lemma 3.1.2 looks at the endpoints of minimal j-connecting paths; Lemma 3.1.3 examines undirected edges, Lemma 3.1.4 and Corollary 3.1.1 look at non-colliders along the path. Lemma 3.2.1 and Corollary 3.2.1 summarize the above lemmas and show that if there is an edge connecting two non-consecutive vertices along a minimal j-connecting path given Z, then the edge is singly directed out of a collider that is in Z, and this edge is real. I first show that the presence of an arrowhead at a vertex on a j-connecting path given Z implies that the vertex is an ancestor of either Z or an endpoint.

Lemma 3.1.1 *If π is a j-connecting path given Z in some joined graph $\mathcal{H}$ and there is an arrowhead at γ_0 on π then γ_0 is either an ancestor of Z or an ancestor of an endpoint in $\mathcal{H}$.*

Proof:
Let α and β be the endpoints of the path π. For a contradiction, assume that π is of the form $\alpha \ldots ? \rightarrow \gamma_0 \ldots \beta$ and that γ_0 is not an ancestor of either Z or an endpoint.

$\beta \neq \gamma_0$ because every vertex is an ancestor of itself. Similarly, γ_0 is not a collider since if $? \rightarrow \gamma_0 \leftarrow ?$ occurs on π then γ_0 is trivially an ancestor of Z (because in this case, γ_0 is a collider on the path π and since π is j-connecting, all colliders are ancestors of Z). Therefore, $? \rightarrow \gamma_0 - ?$ occurs on π. But since π is j-connecting, $? \rightarrow \gamma_0 \rightarrow$ occurs on π.

Let γ be the first vertex after γ_0 that is not a parent of the next vertex on the path. Then γ_0 is an ancestor γ (see Figure 3.1). By assumption γ_0 is not an ancestor of an endpoint so γ is not an endpoint. Let γ_1 be the first vertex after γ along π. By definition γ is not a parent of γ_0, and because π is j-connecting π is of the form $\alpha \ldots ? \rightarrow \gamma_0 \rightarrow \ldots \rightarrow \gamma \leftarrow ? \gamma_1 \ldots \beta$ (where γ_1 possibly equals β). By the definition of j-connection γ, as a collider along the path, is an ancestor of Z and thus γ_0 is an ancestor of Z, which is a contradiction. $\square$

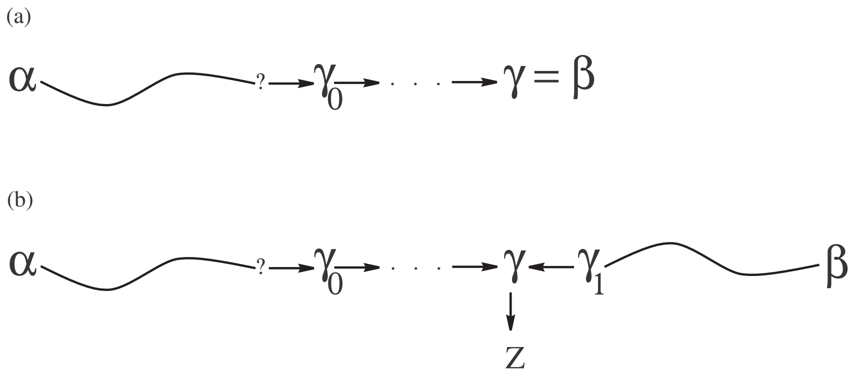


Figure 3.1: Diagram for Proof of Lemma 3.1.1: γ_0 is an ancestor of either an endpoint or Z.

Lemma 3.1.2 *If π is a minimal j-connecting path between α and β given Z in a joined graph $\mathcal{H}$, then if α is adjacent to some non-consecutive (see Definition 2.5.4) vertex γ on the path, then $\alpha \leftarrow \gamma$, γ is a collider on the path π, γ is in Z, and the $\alpha \leftarrow \gamma$ edge is real.*

Proof:

Suppose for a contradiction that there is an edge $\alpha? \rightarrow \gamma^*$ where γ^* is not consecutive to α on π. Let γ be the furthest such vertex from α on π. The path $\alpha? \rightarrow \gamma \dots \beta$ cannot be j-connecting given Z else π is not minimal. There are now three cases to consider:

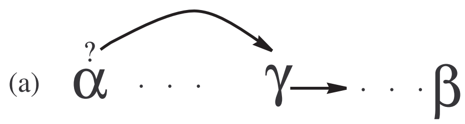


Figure 3.2: Diagram for Proof of Lemma 3.1.2. See text for further details.

(a) If π takes the form $\alpha \dots \gamma \rightarrow \dots \beta$, then $\alpha? \rightarrow \gamma \rightarrow \dots \beta$ forms a shorter j-connecting path since γ is a non-collider on both paths (see Figure 3.2).

(b) If π takes the form $\alpha \dots \gamma \leftarrow ? \dots \beta$ (see Figure 3.3) then by Lemma 3.1.1 γ is either an ancestor of α or an ancestor of a vertex in Z. In the latter case, the path $\alpha? \rightarrow \gamma \leftarrow$

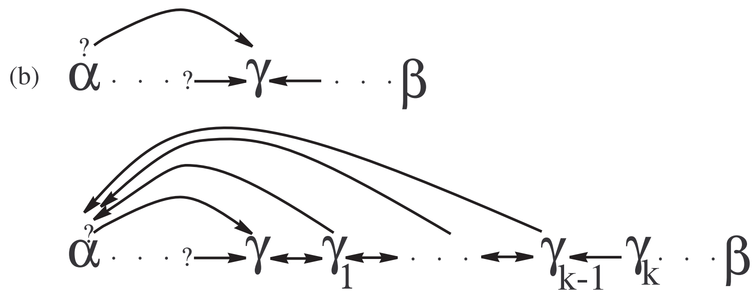


Figure 3.3: Diagram for Proof of Lemma 3.1.2. See text for further details.

$? \ldots \beta$ is j-connecting which is a contradiction. In the former case, $\alpha? \rightarrow \gamma \leftarrow ?\gamma_1$ forms an inducing path hence α and γ_1 are adjacent by Theorem 2.6.1. Since γ is the furthest vertex such that $\alpha? \rightarrow \gamma$ we have $\alpha? - \gamma_1$. Let γ_k be the first vertex after γ_1 which is not adjacent to α (such a vertex exists because otherwise α and β are adjacent). Now $\alpha? - \gamma_{k-1}$ occurs in $\mathcal{H}$ (by the definition of γ). Further, π cannot take the form $\alpha \ldots \gamma_{k-1} - ?\gamma_k \ldots \beta$ since in this case $\alpha? - \gamma_{k-1} - ?\gamma_k \ldots \beta$ is j-connecting which is a contradiction. Hence, π takes the form $\alpha \ldots \gamma_{k-1} \leftarrow ?\gamma_k \ldots \beta$. If $\alpha - \gamma_{k-1}$ then $\alpha - \gamma_{k-1} \leftarrow \gamma_k$, and α and γ_k are adjacent by Lemma 2.4.3 which is a contradiction (by the definition of γ_k). Thus we have $\alpha \leftarrow \gamma_{k-1} \leftarrow ?\gamma_k$. If $\langle \gamma_{k-2}, \gamma_{k-1}, \gamma_k \rangle$ is a non-collider on π, then $\alpha \leftarrow \gamma_{k-1} \leftarrow ?\gamma_k \ldots \beta$ is j-connecting. Hence, γ_{k-1} is a collider on π, and furthermore is in Z. If $k-1 > 1$, $\langle \gamma_k, \gamma_{k-1}, \gamma_{k-2}, \alpha \rangle$ forms a discriminating path for γ_{k-2}. Hence we have $\gamma_{k-2} \rightarrow \alpha$. I can argue that γ_{k-2} is a collider on the path π too and hence if $k-2 > 1$ then $\langle \gamma_k, \gamma_{k-1}, \gamma_{k-2}, \gamma_{k-3}, \alpha \rangle$ forms a discriminating path for γ_{k-3}. Continuing in this way we obtain that $\langle \alpha, \gamma, \gamma_1, \gamma_2, \ldots, \gamma_{k-1}, \gamma_k \rangle$ is a collider sub-path of π such that each of $\langle \gamma_1, \gamma_2, \ldots, \gamma_{k-1} \rangle$ is a parent of α. But now we reach a contradiction: $\langle \alpha, \gamma, \gamma_1, \gamma_2, \ldots, \gamma_{k-1}, \gamma_k \rangle$ forms an inducing path so by Theorem 2.6.1 α and γ_k are adjacent which is a contradiction by the definition of γ_k.

(c) If π takes the form $\alpha \ldots \gamma - \delta \ldots \beta$ then since we have $\alpha? \rightarrow \gamma - \delta$, there is a further edge $\alpha? - ?\delta$ (see Figure 3.4). By definition of γ, $\alpha? - \delta$ occurs in $\mathcal{H}$ but δ is a non-collider (or an endpoint) hence $\alpha? - \delta \ldots \beta$ forms a shorter j-connecting path.

This completes the case argument and establishes that there is no edge $\alpha? \rightarrow \gamma$ in $\mathcal{H}$.

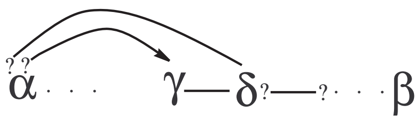
(c) 

Figure 3.4: Diagram for Proof of Lemma 3.1.2. See text for further details.

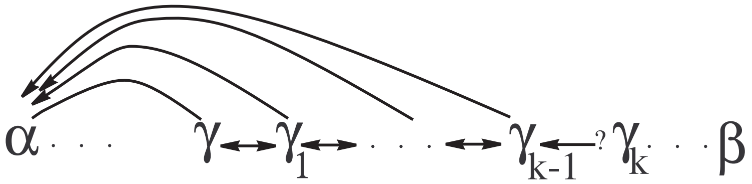


Figure 3.5: Diagram for Proof of Lemma 3.1.2: See text for further details.

Suppose that there is an edge $\alpha - \gamma^*$ where γ^* is not adjacent to α on π. Let γ be the furthest such vertex on π from α (see Figure 3.5). If π takes the form $\alpha \ldots \gamma - ? \ldots \beta$ then $\alpha - \gamma - ? \ldots \beta$ forms a shorter j-connecting path, but if π takes the form $\alpha \ldots \gamma \leftarrow ? \gamma_1 \ldots \beta$ then γ_1 and α are adjacent. Hence, $\alpha \leftarrow \gamma_1$ occurs in $\mathcal{H}$ by definition of γ and the fact that $\alpha ? \rightarrow \gamma_1$ is already ruled out. Let γ_k be the first vertex after γ on π which is not adjacent to α, such exists as otherwise α and β would be adjacent. Hence we have $\alpha \leftarrow \gamma_{k-1} ? - ? \gamma_k$ in $\mathcal{H}$. It follows that $\gamma_j, 1 \leq j \leq k-1$ is a collider on π since otherwise $\alpha \leftarrow \gamma_j \ldots \beta$ forms a shorter j-connecting path. Since each of $\{\gamma, \gamma_1, \ldots, \gamma_{k-1}\}$ are adjacent to, and parents of, α, $\langle \gamma_k, \gamma_{k-1}, \ldots, \gamma, \alpha \rangle$ forms a discriminating path for γ in $\mathcal{H}$, an by Lemma 2.5.4 this discriminating path is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and thus either $\alpha \leftrightarrow \gamma$ or $\alpha \leftarrow \gamma$ is in $\mathcal{H}$, which is a contradiction.

Thus, if α is adjacent to a non-consecutive vertex on π then $\alpha \leftarrow \gamma$. Finally, $\alpha \leftarrow \gamma \ldots \beta$ forms a shorter j-connecting path unless γ is a collider on π and γ is in Z.

To show: $\alpha \leftarrow \gamma$ edge is real.

Suppose for a contradiction that there exists a collider on π that is in Z such that $\alpha \leftarrow \gamma^*$ and this edge is not real. Let γ be the furthest such vertex from α. Let γ_1 be the vertex after γ on π. Since $\langle \gamma, \gamma^*, \gamma_1 \rangle$ is a collider in some $\mathcal{G}$ but not $\mathcal{H}$, α is adjacent to γ_1 where

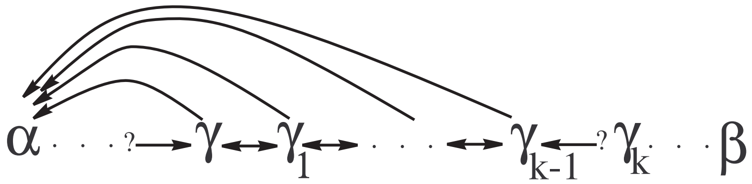


Figure 3.6: Diagram for Proof of Lemma 3.1.2: See text for further details.

γ_1 is the first vertex after γ on π (see Figure 3.5). I have shown that if α is adjacent to γ_1 then $\alpha \leftarrow \gamma_1$ occurs in $\mathcal{H}$, γ_1 is a collider on π and γ_1 is in Z. By the definition of γ, the $\alpha \leftarrow \gamma_1$ edge is real. If γ_2 is not adjacent to α then $\langle \gamma_2, \gamma_1, \gamma, \alpha \rangle$ forms a discriminating path for γ in $\mathcal{H}$ and by Lemma 2.5.4 γ is discriminated to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Since $\gamma \leftrightarrow \gamma_1$ occurs in $\mathcal{H}$, $\gamma \leftrightarrow \gamma_1$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ and consequently, the $\alpha \leftarrow \gamma$ edge is real. But this is a contradiction so γ_2 is adjacent to α. Let γ_k be the first vertex after γ that is not adjacent to α. Such a vertex exists because α and β are not adjacent. I can argue that each of $\{\gamma_1, \gamma_2, \ldots, \gamma_{k-1}\}$ is a collider on π that is in Z, a parent of α in $\mathcal{H}$, and the edges $\alpha \leftarrow \gamma_r, 1 \leq r < k$ occur in $\mathcal{G}$. By the definition of γ, the $\alpha \leftarrow \gamma_r, 1 \leq r < k$ edges are real and so occur in $\mathcal{G}$. But then $\langle \gamma_k, \gamma_{k-1}, \ldots \gamma_1, \gamma, \alpha \rangle$ forms a discriminating path for γ such that the $\alpha \leftarrow \gamma$ edge is real. This is a contradiction, so the $\alpha \leftarrow \gamma$ edge is real. $\square$

Corollary 3.1.1 *If π is a minimal j-connecting path between α and β given Z, and δ is a vertex such that π takes the form $\alpha \ldots \leftarrow \delta \ldots \beta$, then if δ is adjacent to a non-consecutive vertex γ occurring after δ on the path, then $\delta \leftarrow \gamma$ and γ is a collider on π, γ is in Z and the $\delta \leftarrow \gamma$ edge is real.*

Proof:

This follows by the same argument as for Lemma 3.1.2, after observing that if π is j-connecting given Z then $\delta \notin Z$, so $\pi(\delta, \beta)$ j-connects given Z. Further, if δ is adjacent to γ and $\delta ? - ? \gamma \ldots \beta$ is j-connecting α and β given Z then $\alpha \ldots \delta ? - ? \gamma \ldots \beta$ is also j-connecting given Z.

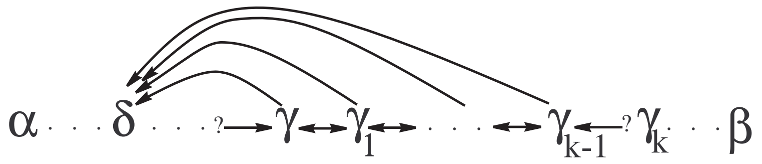


Figure 3.7: Diagram for Proof of Lemma 3.1.1: See text for further details.

To show: $\delta \leftarrow \gamma$ edge is real. Suppose for a contradiction that there exists a collider on π that is in Z such that $\delta \leftarrow \gamma^*$ and this edge is not real. Let γ be the furthest such vertex from δ (see Figure 3.7). Then δ is adjacent to γ_1 where γ_1 is the first vertex after γ on π. I have shown that if δ is adjacent to γ_1 then $\delta \leftarrow \gamma_1$ occurs in $\mathcal{H}$, γ_1 is a collider on π and γ_1 is in Z. By the definition of γ, the $\delta \leftarrow \gamma_1$ edge is real. Let γ_2 be the vertex after γ_1 on π. If γ_2 is not adjacent to δ then $\langle \gamma_2, \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ which would imply that the $\delta \leftarrow \gamma$ edge is real. But this is a contradiction so γ_2 is adjacent to δ. Let γ_k be the first vertex after γ that is not adjacent to δ (such a vertex exists because by Lemma 3.1.2 δ is not adjacent to β). We can argue that each of $\{\gamma_1, \gamma_2, \ldots, \gamma_{k-1}\}$ is a collider on π that is in Z, a parent of δ in $\mathcal{H}$, and the edges $\delta \leftarrow \gamma_r, 1 \leq r < k$ occur in $\mathcal{G}$. By the definition of γ, the $\delta \leftarrow \gamma_r, 1 \leq r < k$ edges are real. But then $\langle \gamma_k, \gamma_{k-1}, \ldots \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ such that the $\delta \leftarrow \gamma$ edge is real. This is a contradiction, so the $\delta \leftarrow \gamma$ edge is real. $\square$

Lemma 3.1.3 *Let π be a minimal j-connecting path between α and β given Z, and δ be a vertex such that π takes the form $\alpha \ldots - \delta \ldots \beta$. If δ is adjacent to a non-consecutive vertex γ occurring after δ on the path, then $\delta \leftarrow \gamma$ and γ is a collider on π, γ is in Z, and the $\delta \leftarrow \gamma$ edge in $\mathcal{H}$ is real.*

Proof:

There are four steps to this proof:

(1) **Establish that γ is of the form $\gamma \leftarrow ? \ldots \beta$ on π.**

Suppose for a contradiction that there exists a γ^* that is a vertex occurring after δ on π and non-consecutive to δ such that γ^* is adjacent to δ and γ^* does not take the form $\gamma^* \leftarrow ? \ldots \beta$ along π. Since δ is not a collider on π, by Lemma 3.1.2 γ^* is not an endpoint and by Corollary 3.1.1 γ^* does not take the form $\gamma^* \rightarrow \ldots \beta$ on π. Then by process of elimination $\gamma^* - \ldots \beta$ is in $\mathcal{H}$.

(1)

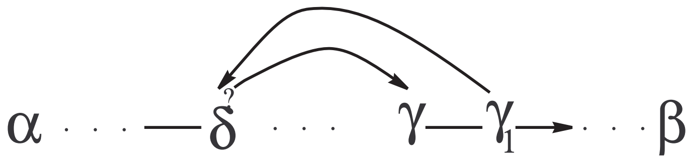


Figure 3.8: Diagram for Proof of Lemma 3.1.3: See text for further details.

Choose γ to be the vertex furthest from δ that is like γ^* (see Figure 3.8). I will now show that in this case, each type of edge between δ and γ leads to a contradiction. If $\delta - \gamma$ is in $\mathcal{H}$, then the path $\alpha \ldots - \delta - \gamma - \ldots \beta$ is a j-connecting path between α and β that is more minimal than π, which is a contradiction. Suppose $\delta ? \rightarrow \gamma$ is in $\mathcal{H}$. Let γ_1 be the vertex after γ on π. Then by Lemma 2.5.1, the $\gamma - \gamma_1$ edge (where γ_1 is the first vertex after γ along π) is not real and δ is adjacent to γ_1. By the definition of γ, $\gamma_1 - \ldots \beta$ is not in $\mathcal{H}$, and because π is j-connecting, $\gamma - \gamma_1 \leftarrow ? \ldots \beta$ does not occur on π in $\mathcal{H}$. Hence γ_1 is either an endpoint or takes the form $\gamma - \gamma_1 \rightarrow \ldots \beta$ on π. But if δ and γ_1 are adjacent, then by Lemma 3.1.2 and Corollary 3.1.1 respectively, δ would be a collider on π which is a contradiction. Therefore, if δ is adjacent to γ, γ is of the form $\gamma \leftarrow ? \ldots \beta$ along π.

(2) **Establish that $\delta \leftarrow \gamma$ is in $\mathcal{H}$.**

Suppose there exists a γ^* that is non-consecutive to δ^* on π such that γ^* is adjacent to δ and γ^* takes the form $\delta ? - ? \gamma^* \leftarrow ? \ldots \beta$ on π. Choose δ and γ to be the vertices furthest apart along π such that δ and γ satisfy the conditions on δ^* and γ^* respectively.

There are 3 sub-cases to consider:

(i) Suppose $\delta \rightarrow \gamma$ is in $\mathcal{H}$ (see Figure 3.9(2i)). Then $- \delta \rightarrow \gamma \leftarrow ?$ is j-connecting

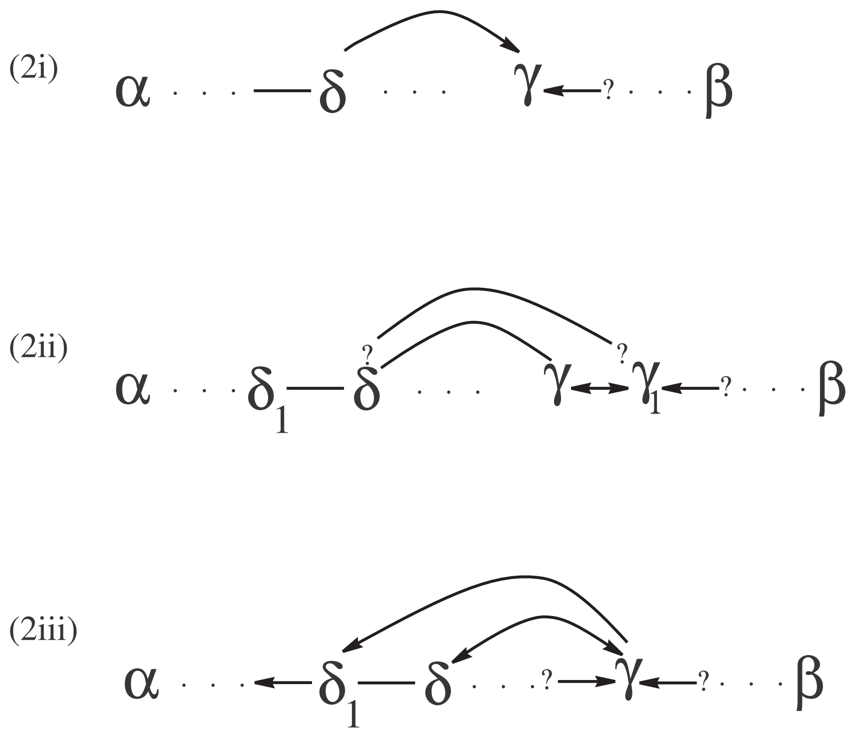


Figure 3.9: Diagram for Proof of Lemma 3.1.3: See text for further details.

given Z because by Lemma 3.1.1 γ is an ancestor of Z. Since the path $\langle \alpha \ldots \delta, \gamma, \ldots \beta \rangle$ violates the minimality of π, we reach a contradiction and $\delta \rightarrow \gamma$ is not in $\mathcal{H}$.

(ii) Suppose $-\delta - \gamma \leftarrow ?$ is in $\mathcal{H}$ (see Figure 3.9(2ii)). Then by Lemma 2.5.1 the $\delta - \gamma$ edge is not real and δ is adjacent to γ_1 where γ_1 is the first vertex after γ along π.

Because δ is of the form $\alpha \ldots - \delta$ along π, by Lemma 3.1.2 and Corollary 3.1.1 respectively, γ_1 is neither an endpoint nor a non-collider of the form $\gamma \leftarrow ? \gamma_1 \rightarrow \ldots \beta$ along π. By (1) I have shown that the edge after γ_1 is not undirected along π. Hence γ_1 is of the form $\gamma_1 \leftarrow ? \ldots \beta$ along π; but this is a contradiction by the definition of γ (since δ and γ are non-consecutive vertices furthest apart on π such that $\alpha \ldots - \delta \ldots \gamma \leftarrow ? \ldots \beta$ is in $\mathcal{H}$ and δ is adjacent to γ). Thus, $\delta - \gamma$ is not in $\mathcal{H}$.

(iii) Suppose $\delta \leftrightarrow \gamma$ is in $\mathcal{H}$ (see Figure 3.9(2iii)). Since we have an arrowhead meeting an undirected edge ($-\delta \leftrightarrow \gamma$), by Lemma 2.4.3 γ is adjacent to δ_1, where δ_1 is the vertex directly before δ on the path π. Since π is j-connecting given Z, the configuration $? \rightarrow \delta_1 - \delta$ does not occur on π. By the definition of the $\{\delta, \gamma\}$ edge, the configuration $-\delta_1 - \delta$ does not occur on π; hence δ_1 is either an endpoint or a non-collider of the form $\leftarrow \delta_1 - \delta$. Consequently, by Lemma 3.1.2 and Corollary 3.1.1 respectively, $\delta_1 \leftarrow \gamma$ is in $\mathcal{H}$ and this edge is real. But here we reach a contradiction: by Lemma 2.5.2 and because the configuration $\delta \leftrightarrow \gamma \rightarrow \delta_1 ?-? \delta$ is in $\mathcal{H}$ with the $\gamma \rightarrow \delta_1$ edge real, the $\{\delta_1, \delta\}$ edge is not undirected, which is contradiction.

I have shown by process of elimination that if δ and γ are adjacent, then $\delta \leftarrow \gamma$ occurs in $\mathcal{H}$.

(3) **γ is a collider on π and γ is in Z.**

(3)

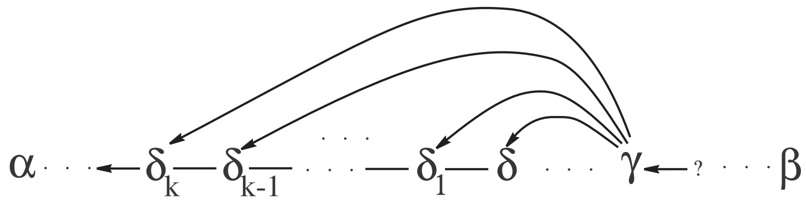


Figure 3.10: Diagram for Proof of Lemma 3.1.3: See text for further details.

Suppose for a contradiction that γ is not a collider, and is thus not in Z (see Figure 3.10). Since $-\delta \leftarrow \gamma$ is in $\mathcal{H}$, γ is adjacent to δ_1 where δ_1 is the vertex directly before δ on π. If δ_1 is an endpoint, then by Lemma 3.1.2 γ is a collider on π and γ is in Z, which is a contradiction. If δ_1 is not an endpoint, then since π is j-connecting, δ_1 is of the form $?-\delta_1-\delta$ on π. Let δ_k be the vertex closest to δ that is before δ and is not a neighbour of its predecessor on the path (or is an endpoint) δ. By repeated applications of the undirected-edge-meeting-an-arrowhead argument, it can be shown that each of $\langle \delta_k, \delta_{k-1}, \ldots \delta_1 \rangle$ is adjacent to γ and by (2) these edges are directed from γ to each of $\langle \delta_{k-1}, \delta_{k-2}, \ldots, \delta_1 \rangle$. If δ_k is an endpoint, then by Lemma 3.1.2 γ is a collider on π and γ is in Z, which is a contradiction. If δ_k is not an endpoint, then $\leftarrow \delta_k - \delta_{k-1}$ is in $\mathcal{H}$ because π is j-connecting. In this case, I can apply Corollary

3.1.1 to see that γ is a collider on π and γ is in Z, which is a contradiction. Therefore, I have shown that if $\delta \leftarrow \gamma$ occurs in $\mathcal{H}$, then γ is a collider on π and γ is in Z.

(4) **Show the $\delta \leftarrow \gamma$ edge is real.**

Suppose δ^* and γ^* are non-consecutive vertices on π such that π is of the form $\alpha \ldots - \delta^* \ldots ? \rightarrow \gamma^* \leftarrow ? \ldots \beta$, $\delta^* \leftarrow \gamma^*$, and γ^* is a collider on π and is in Z. Suppose for a contradiction that the $\delta^* \leftarrow \gamma^*$ edge is not real. Choose δ and γ to be the vertices that are furthest apart along π and satisfy the conditions on δ^* and γ^* respectively (see Figure 3.11).

(4)

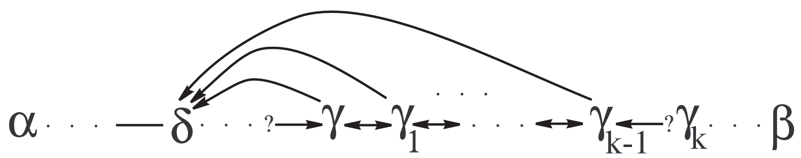


Figure 3.11: Diagram for Proof of Lemma 3.1.3: See text for further details.

Because the $\delta \leftarrow \gamma$ edge is not real, δ is adjacent to γ_1 where γ_1 is the first vertex after γ on π. I have shown that if δ is adjacent to γ_1 then $\delta \leftarrow \gamma_1$ occurs in $\mathcal{H}$, γ_1 is a collider on π and γ_1 is in Z. By the definition of γ, the $\delta \leftarrow \gamma_1$ edge is real. If γ_2 is not adjacent to δ then $\langle \gamma_2, \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ which would imply that the $\delta \leftarrow \gamma$ edge is real. But by Lemma 2.1.1 this is a contradiction so γ_2 is adjacent to δ. Let γ_k be the first vertex after γ that is not adjacent to δ. Such a vertex exists because by Lemma 3.1.2 β is not adjacent to δ since δ is not a collider on π. I can argue that each of $\{\gamma_1, \gamma_2, \ldots, \gamma_{k-1}\}$ is a collider on π that is in Z, a parent of δ in $\mathcal{H}$. But then $\langle \gamma_k, \gamma_{k-1}, \ldots \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ in $\mathcal{H}$ so by Lemma 2.1.1 the $\delta \leftarrow \gamma$ edge is real. This is a contradiction, so the $\delta \leftarrow \gamma$ edge is real. □

Lemma 3.1.4 *Let π be a minimal j-connecting path between α and β given Z in a joined graph $\mathcal{H}$ formed by joining a number of Markov equivalent maximal ancestral graphs. If π is of the form $\alpha \ldots ? \rightarrow \delta \ldots \gamma \leftarrow ? \ldots \beta$ and $\delta ? - \gamma$ occurs in $\mathcal{H}$, then $\delta \leftarrow \gamma$, γ is a collider on π, γ is in Z, and the $\delta \leftarrow \gamma$ edge in $\mathcal{H}$ is real.*

If π is of the form $\alpha \ldots ? \rightarrow \delta \ldots \gamma \leftarrow ? \ldots \beta$ and δ and γ are adjacent, then the $\{\delta, \gamma\}$ edge cannot be bi-directed or else the minimality of the path π would be violated. I now show that in such situations, there is a directed edge, starting from a collider and this directed edge is real. The following proof is very similar to the proof of Lemma 2.5.5.

Proof:

Suppose, for a contradiction, that there is an edge $\delta^* ? - \gamma^*$ in $\mathcal{H}$, and the tail at γ^* is not real. Let δ and γ be the vertices furthest apart along π such that δ and γ satisfy the conditions on δ^* and γ^* respectively. Let γ_1 be the vertex just after γ along π in the direction of β; by hypothesis $\gamma \leftarrow ? \gamma_1$.

1. **To show $\{\delta, \gamma_1\}$ are adjacent.**
   If $\{\delta, \gamma_1\}$ are not adjacent, there exists at least one $\mathcal{G}$ in which $\langle \delta, \gamma, \gamma_1 \rangle$ forms an unshielded non-collider since it is an unshielded non-collider in $\mathcal{H}$. Because all $\mathcal{G}$ being joined are Markov equivalent, $\langle \delta, \gamma, \gamma_1 \rangle$ is an unshielded non-collider in all $\mathcal{G}$. Since ancestral graphs do not contain undirected edges that meet arrowheads, $\gamma \leftarrow ? \gamma_1$ is in all $\mathcal{G}$, and the tail at γ is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Consequently: $\delta \leftarrow \gamma$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ so $\delta \leftarrow \gamma$ occurs in $\mathcal{H}$ and this edge is real. But this is a contradiction because we are assuming that there is at least one graph $\mathcal{G}$, that gave rise to $\mathcal{H}$, without $\delta \leftarrow \gamma$. So, $\{\delta, \gamma_1)$ are adjacent in $\mathcal{H}$.
2. **Establish that $\delta \leftarrow \gamma_1$ in $\mathcal{H}$ (see Figure 3.12).**

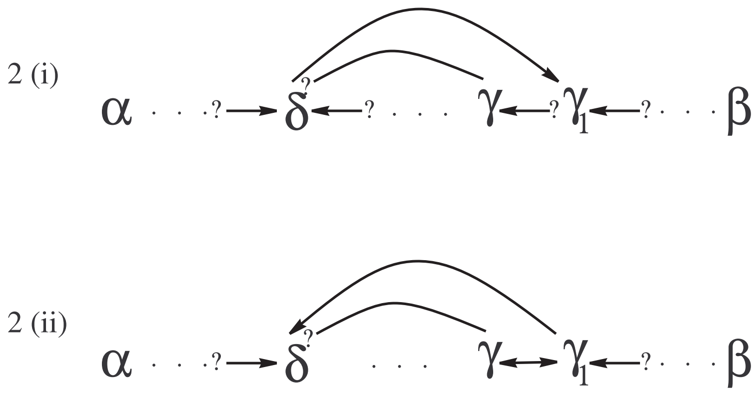


Figure 3.12: Diagram for Proof of Lemma 3.1.4, step 2.

If γ_1 is an endpoint or a non-collider of the form $\gamma \leftarrow ?\gamma_1 - ?$, then by Lemma 3.1.2, Lemma 3.1.3 and Corollary 3.1.1 $\delta \rightarrow \gamma_1$, δ is a collider on π and a vertex in Z, and the $\delta \rightarrow \gamma_1$ edge is real. But then we reach a contradiction: $\delta \rightarrow \gamma_1 ? \rightarrow \gamma - ?\delta$ is in $\mathcal{H}$ and by Lemma 2.5.2 there is an arrowhead at γ on the $\{\delta, \gamma\}$ edge in $\mathcal{H}$, contrary to hypothesis. Therefore, γ_1 satisfies the conditions on γ and $\gamma \leftarrow ?\gamma_1 \leftarrow ?$ is in $\mathcal{H}$.

By the definition of δ and γ, the $\{\delta, \gamma_1\}$ edge in $\mathcal{H}$ is real. If $\delta \rightarrow \gamma_1$ is in $\mathcal{H}$ then since this edge is real, we reach a contradiction because by Lemma 2.5.2 there is an arrowhead at γ on the $\{\delta, \gamma\}$ edge contrary to hypothesis. Therefore $\delta \leftarrow \gamma_1$ is in $\mathcal{H}$ and this edge is real. Furthermore, if γ_1 is not a collider on π and hence is not a vertex in Z, then the path $\langle \alpha, \ldots, \delta, \gamma_1, \ldots, \beta \rangle$ is j-connecting given Z and is more minimal than π. Hence, $\delta \leftarrow \gamma_1$ is in $\mathcal{H}$, this edge is real, γ_1 is a collider on π, and γ_1 is a vertex in Z.

3. **Establish that $\delta \leftarrow \gamma$ in $\mathcal{H}$.**

   Suppose for a contradiction that $\delta - \gamma$ in $\mathcal{H}$. Let γ_2 be the vertex after γ_1 on π. If γ_2 is not adjacent to δ then $\langle \gamma_2, \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ and consequently, $\delta \leftarrow \gamma$ would occur in $\mathcal{H}$ which is a contradiction. Therefore, γ_2 is adjacent to δ in $\mathcal{H}$. Since $\gamma_1 \rightarrow \delta$ is real, by Lemma 2.5.2 and minimality $\delta \leftarrow \gamma_2$, and by the definition of δ and γ, this edge is real. Furthermore, since π is a minimal j-connecting path given Z, γ_2 is a collider on π, and γ_2 is a vertex in Z. Let γ_k be the first vertex after γ_1 that is not a collider in Z. It can be verified that each of $\{\gamma_1, \gamma_2, \ldots, \gamma_{k-1}\}$ is a parent of δ and these $\delta \leftarrow \gamma_i, 1 \leq i \leq (k-1)$ edges are real. There are three cases to be considered (see Figure 3.13):

   i) If γ_k and δ are not adjacent then $\langle \gamma_k, \gamma_{k-1}, \ldots, \gamma_1, \gamma, \delta \rangle$ forms a discriminating path for γ in $\mathcal{H}$, and by Lemma 2.5.4 $\delta \leftarrow \gamma$ is in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. But this is a contradiction because we are assuming that the $\delta \leftarrow \gamma$ edge in $\mathcal{H}$ is not real.

   ii) Suppose γ_k and δ are adjacent and γ_k is of the form $\gamma_{k-1} \leftarrow ?\gamma_k \leftarrow ?$. By the definition of γ_k, γ_k is not a vertex in Z. By Lemma 2.5.2 $\delta \leftarrow \gamma_k$ is in $\mathcal{H}$, and by the definition of δ and γ, this edge is real. Furthermore, since γ_k is not in Z, $\langle \alpha, \ldots, \delta, \gamma_k, \ldots \beta \rangle$ forms a more minimal j-connecting path given Z than π, which is a contradiction.

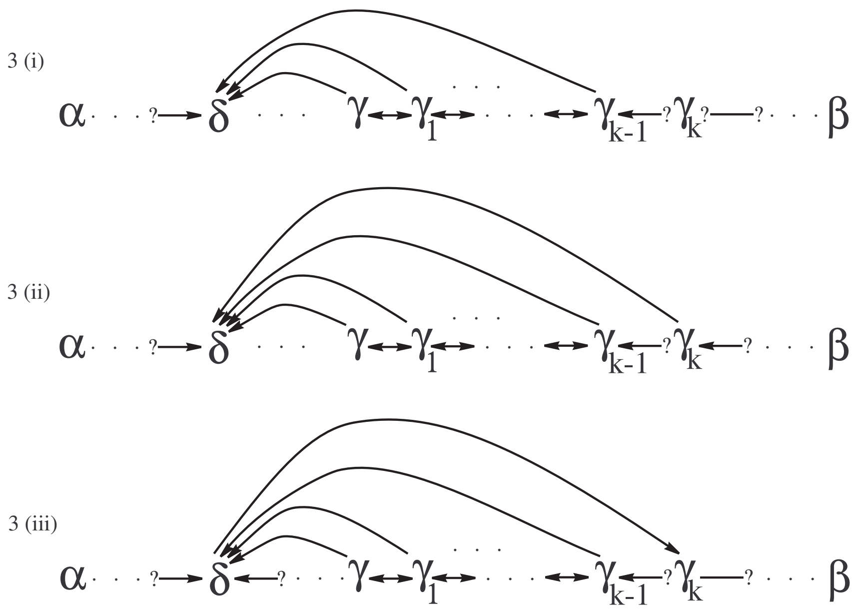


Figure 3.13: Diagram for Proof of Lemma 3.1.4, step 3.

iii) Suppose γ_k and δ are adjacent and γ_k is either an endpoint or is of the form $\gamma_{k-1} \leftarrow ?\gamma_k - ?$. Then $\delta \to \gamma_k$ is in $\mathcal{H}$, δ is a collider on π, δ is in Z and the $\delta \to \gamma_k$ edge is real (by Lemma 3.1.2 in the former case; and by Corollary 3.1.1 and Lemma 3.1.3 in the latter case). Because the $\delta \leftarrow \gamma_{k-1}$ and $\delta \to \gamma_k$ edges are real in $\mathcal{H}$, the non-ancestral configuration $\delta \to \gamma_k ? \to \gamma_{k-1} \to \delta$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ which is a contradiction.

I have shown that if $\delta ? - \gamma$ is in $\mathcal{H}$, then $\delta \leftarrow \gamma$ is in $\mathcal{H}$ and this edge is real. Furthermore, γ is a collider on π and is a vertex in Z or else the path $\langle \alpha, \ldots, \delta, \gamma, \ldots, \beta \rangle$ forms a j-connecting path between α and β given Z that is more minimal than π, which is a contradiction. □

Corollary 3.1.2 *Let π be a minimal j-connecting path between α and β given Z. If ν and κ are non-consecutive colliders on π that are adjacent, then: (i) either $\nu \to \kappa$ or $\nu \leftarrow \kappa$ occurs in $\mathcal{H}$, (ii) the $\{\nu, \kappa\}$ edge is real, and (iii) the parent of the $\{\nu, \kappa\}$ edge is in Z.*

Proof:

Suppose without loss of generality that $\pi = \langle \alpha, \ldots, \nu, \ldots, \kappa, \ldots, \beta \rangle$. Since ν and κ are colliders, π is of the form $\alpha \ldots ? \rightarrow \nu \ldots \kappa \leftarrow ? \ldots \beta$. By the minimality of the path π, the $\{\nu, \kappa\}$ edge is not bi-directed. There are two cases to consider: either $\nu - ?\kappa$ is in $\mathcal{H}$, or $\nu ? - \kappa$ is on $\mathcal{H}$. In the former case, $\nu \rightarrow \kappa$ is in $\mathcal{H}$, this edge is real and ν is in Z by Lemma 3.1.4. A symmetric argument can be made for the latter case. □

3.2 Structure of j-connecting Paths

The following two lemmas summarize the results thus far and characterize the structure of minimal j-connecting paths given Z. Because DAGs are a subset of joined graphs, and j-connection for DAGs is equivalent to d-connection, these results are generalizable to minimal d-connecting paths given Z. Similarly, j-connection for maximal ancestral graphs is equivalent to m-connection, so these results also characterize minimal m-connecting paths for maximal ancestral graphs.

Lemma 3.2.1 *Let π be a minimal j-connecting path between α and β given Z and let δ be either an endpoint or a non-collider along π. If δ is adjacent to γ, a non-consecutive vertex along π, then $\delta \leftarrow \gamma$ is in $\mathcal{H}$, γ is a collider on π, γ is in Z and the $\delta \leftarrow \gamma$ edge is real.*

Proof:

Claim: γ is a collider on π, γ is in Z and the $\delta \leftarrow \gamma$ edge is real.

There are four cases to consider:

(a) If δ is an endpoint then by Lemma 3.1.2 the claim is true.

(b) If π is of the form $\alpha \ldots \leftarrow \delta \ldots \gamma \ldots \beta$ then by Corollary 3.1.1 the claim is true.

(c) If π is of the form $\alpha \ldots - \delta \ldots \gamma \ldots \beta$ then by Lemma 3.1.3 the claim is true.

(d) If π is of the form $\alpha \ldots ? \rightarrow \delta \rightarrow \ldots \gamma \ldots \beta$ then by Corollary 3.1.1 and Lemma 3.1.3 γ is not an endpoint and $\gamma - ? \ldots \beta$ cannot occur on π so $\gamma \leftarrow ? \ldots \beta$ is in $\mathcal{H}$. By Lemma 3.1.4 either $\delta \rightarrow \gamma$ or $\delta \leftarrow \gamma$ is in $\mathcal{H}$. $\delta \rightarrow \gamma$ is ruled out as in this case δ is a collider on π. Hence $\delta \leftarrow \gamma$ on π in $\mathcal{H}$ and the claim is true. □

Corollary 3.2.1 *Let π be a minimal j-connecting path between α and β given Z; let δ and γ be two non-consecutive vertices on π that are adjacent. Then at least one of δ and γ is a collider in Z, the $\{\delta, \gamma\}$ edge is directed out of a collider in Z and this edge is real.*

Proof:

There are two cases to consider:

(i) If one of δ or γ is a non-collider or an endpoint (say δ), then by Lemma 3.2.1 the other vertex γ is a collider in Z, the $\{\delta, \gamma\}$ edge is directed out of γ and this edge is real.

(ii) If both δ and γ are colliders, then by Corollary 3.1.2 the $\{\delta, \gamma\}$ edge is directed, the parent of the $\{\delta, \gamma\}$ edge is in Z and this edge is real. $\Box$

If π is any minimal j-connecting path given Z, then π can be decomposed into a series of undirected, directed and bi-directed sub-paths. Furthermore, if π contains both colliders and undirected edges, then there is at least one edge on π that is oriented from the undirected sub-path to the directed sub-path. Figure 3.14 shows an example of an induced subgraph for a j-connecting path given Z.

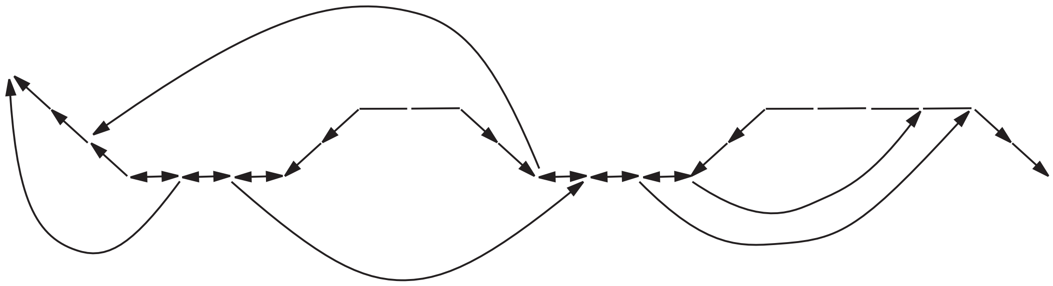


Figure 3.14: Example of a minimal j-connecting path given Z in a joined graph

Thus, it is apparent that the minimality of a j-connecting path π puts a number of structural restrictions on the induced subgraph formed by the vertices on the path π. This observation will facilitate the proof of the global Markov property for joined graphs.

3.3 Global Markov Property for Joined Graphs

The global Markov property for joined graphs is defined as follows:

$$A \perp\!\!\!\perp B | Z[P] \text{ whenever } Z \text{ j-separates } A \text{ and } B \text{ in a joined graph.}$$

where A, B and Z are disjoint subsets of vertices in the joined graph. The sets of vertices A and B are j-separated given the set Z if and only if every vertex in A is j-separated from every vertex in B given Z. In other words, if there is at least one j-connecting path conditional on Z from some vertex in A to some vertex in B, then A and B are j-connecting given Z.

Theorem 3.8.1 proves that the global Markov property holds for P if the joined graph is formed by joining a set of Markov equivalent maximal ancestral graphs and P obeys the global Markov property with respect to the graphs. There are two parts to this proof:

- j-connection given Z in the joined graph $\Rightarrow$ m-connection given Z in the set of graphs joined to give rise to the joined graph.

- m-connection given Z in the set of graphs joined to give rise to the joined graph $\Rightarrow$ j-connection given Z in the joined graph.

Note that m-connection is the same as j-connection for ancestral graphs because arrowheads do not meet undirected edges in ancestral graphs. In particular, I look at minimal j-connecting paths given Z in the joined graph, and show that there exists some j-connecting path between the endpoints given Z in the ancestral graphs.

Figure 3.15 can be thought of as a "road map"for proving Theorem 3.8.1. The proofs of Lemmas 3.5.2, 3.5.3, 3.5.4, 3.5.5 and 3.5.6 are very similar to proving the maximality result (in particular Lemmas 2.6.2, 2.6.3, 2.6.4, 2.6.5 and Theorem 2.6.1). Lemma 3.2.1 and Corollary 3.2.1 describe the structure of minimal j-connecting paths given Z. Lemma 3.7.1 and Corollaries 3.7.1, 3.7.2, and 3.7.3 are required to ensure that *discriminating paths with order* are well-defined (see Definition 3.7.1). Discriminating paths with order are used in Lemma 3.7.3, hence the dashed line. Lemma 3.7.3 can then be used to show that minimal j-connecting paths given Z are retained through the join operation. Theorem A.0.1 states conditions under which two maximal ancestral graphs are Markov equivalent to each other.

As a consequence of the structure of minimal j-connecting paths, all non-colliders in a joined graph $\mathcal{H}$, formed by joining a set of Markov equivalent maximal ancestral graphs,

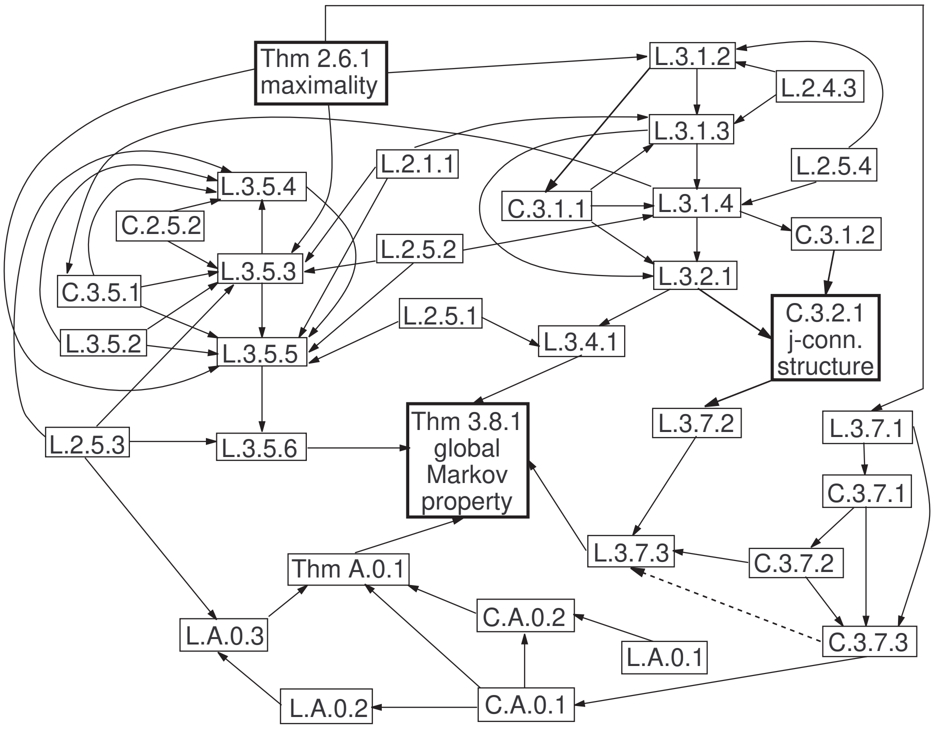


Figure 3.15: Road Map for J-connection Proof. See text for details.

are also non-colliders in all ancestral graphs that gave rise to $\mathcal{H}$. I prove this property in the next section.

3.4 Non-Colliders on Minimal j-connecting Paths in Joined Graphs

Lemma 3.4.1 *Let π be a minimal j-connecting path between α and β given Z in some joined graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs. Then, all non-colliders on π in $\mathcal{H}$ are also non-colliders on π in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.*

Proof:

Suppose for a contradiction that $\langle \delta, \rho, \gamma \rangle$ is a non-collider on π in $\mathcal{H}$ that is not present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Then by Lemma 2.5.1 at least one of $\{\delta, \rho\}$ and $\{\rho, \gamma\}$ is not real, δ is adjacent to γ and by Corollary 3.2.1 this edge is directed. There are three cases to consider:

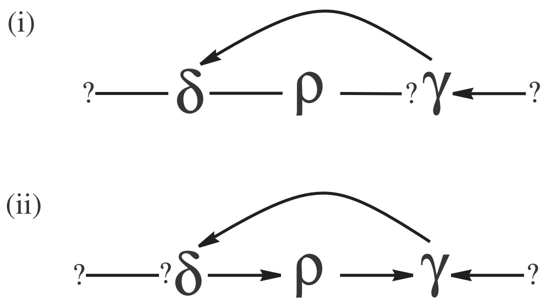


Figure 3.16: Diagram for Proof of Lemma 3.4.1.

(i) $\delta - \rho$ is in $\mathcal{H}$ (see Figure 3.16(i)). Since π is j-connecting, δ is either an endpoint or a non-collider on π. Then by Lemma 3.2.1 γ is a collider on π, γ is in Z, $\delta \leftarrow \gamma$ is in $\mathcal{H}$ and this edge is real. But then $\gamma \rightarrow \delta -?\rho \rightarrow \gamma$ is in $\mathcal{H}$, and by Lemma 2.5.2 we can infer an arrowhead at ρ on the $\{\delta, \rho\}$ edge which is a contradiction. Therefore $\delta - \rho$ does not occur in $\mathcal{H}$.

(ii) $\delta \rightarrow \rho$ is in $\mathcal{H}$ (see Figure 3.16(ii)). Since π is j-connecting, δ is either an endpoint or a non-collider on π. Then by Lemma 3.2.1 γ is a collider on π, γ is in Z, $\delta \leftarrow \gamma$ is in $\mathcal{H}$ and this edge is real. But then $\gamma \rightarrow \delta \rightarrow \rho \rightarrow \gamma$ is in $\mathcal{H}$, and by Lemma 2.5.2 we can infer an arrowhead at ρ on the $\{\delta, \rho\}$ edge which is a contradiction. Therefore $\delta \rightarrow \rho$ does not occur in $\mathcal{H}$.

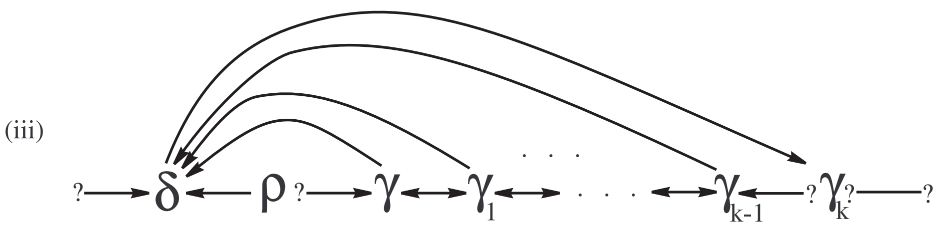


Figure 3.17: Diagram for Proof of Lemma 3.4.1. See text for details.

(iii) $\delta \leftarrow \rho \rightarrow \gamma$ or $\gamma \leftrightarrow \rho \rightarrow \gamma$ or $\delta \leftarrow \rho \leftrightarrow$ is in $\mathcal{H}$ (see Figure 3.17(iii)). Without loss of generality, suppose $\delta \leftarrow \gamma$ is in $\mathcal{H}$. Then by Lemma 3.1.4 γ is a collider in $\mathcal{H}$, γ

is in Z, and the $\delta \leftarrow \gamma$ edge is real. Let γ_1 be the vertex just after γ on π. If γ_1 is not adjacent to δ then $\langle \gamma_1, \gamma, \rho, \delta \rangle$ forms a discriminating path for ρ and by Lemma 2.5.4 this path is discriminating in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and hence the collider $\langle \gamma, \rho, \delta \rangle$ exists in all $\mathcal{G}$, which is a contradiction. Therefore γ_1 is adjacent to δ. By Lemma 2.5.2, $\delta \leftarrow \gamma_1$ and by Lemma 3.1.4 γ_1 is a collider on π, γ_1 is in Z and the $\delta \leftarrow \gamma_1$ edge is real. Let γ_k be the first vertex after γ on π that is not a collider on π and a parent of δ. Then each of $\langle \gamma_1, \ldots, \gamma_{k-1} \rangle$ is a collider on π and a parent of δ. Furthermore by Corollary 3.2.1 these $\delta \leftarrow \gamma_r, 1 \leq r < k$ edges are real. If γ_k is either an endpoint or a non-collider, then by Lemma 3.2.1 δ is a collider on π, $\delta \rightarrow \gamma_k$ and this edge is real. But this is a contradiction because $\delta \rightarrow \gamma_k ? \rightarrow \gamma_{k-1} \rightarrow \delta$ would occur in any $\mathcal{G}$ that gave rise to $\mathcal{H}$ and these graphs would not be ancestral. But if δ is not adjacent to γ_k then $\langle \gamma_k, \gamma_{k-1}, \ldots, \gamma_1, \gamma, \rho, \delta \rangle$ forms a discriminating path for ρ which is a contradiction.

Therefore, I have shown that if $\langle \delta, \rho, \gamma \rangle$ forms a non-collider on a minimal j-connecting path given Z in $\mathcal{H}$, then $\langle \delta, \rho, \gamma \rangle$ is a non-collider on π in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

3.5 Collider Sub-paths of Minimal j-connecting Paths in Joined Graphs

This section proves that collider sub-paths of minimal j-connecting paths in a joined graph $\mathcal{H}$, formed by joining Markov equivalent maximal ancestral graphs, are also m-connecting in the ancestral graphs joined to give rise to $\mathcal{H}$. The steps of this proof are very similar to the proof that $\mathcal{H}$ is maximal (Theorem 2.6.1). Essentially, the claims made about which non-consecutive edges are adjacent, and the orientation of such edges are the same, but the arguments to justify these claims differ slightly.

Having established that the graph resulting from joining maximal ancestral graphs is itself maximal, I can prove the following lemma, which is analogous to Lemma 2.6.1 for j-connection:

Lemma 3.5.1 *Let $\mathcal{H}$ be a graph resulting from joining a set of Markov equivalent maximal ancestral graphs and π is a minimal j-connecting path between α and β given Z in $\mathcal{H}$.*

Furthermore, let μ be a collider sub-path of π with endpoints μ_0 and $\mu_n, n > 2$ such that neither μ_0 nor μ_n are colliders on π. If such a path μ exists in $\mathcal{H}$, then the configurations $\mu_j \to \alpha_0 \to \mu_{j+1}$ and $\mu_{j-1} \leftarrow \alpha_0 \leftarrow \mu_j$ (provided these vertices exist) do not occur in $\mathcal{H}$ for $1 \leq j \leq (n-1)$ where α_0 is a child of μ_j.

Proof:

Note that by Lemma 3.2.1 μ_0 and μ_n are not adjacent, as neither of these vertices is a collider. For a contradiction, suppose that $\mu_j \to \alpha_0 \to \mu_{j+1}$ occurs in $\mathcal{H}$. Then $\langle \mu_{j-1}, \mu_j, \mu_{j+1} \rangle$ forms an inducing path in $\mathcal{H}$ so by Theorem 2.6.1 μ_{j-1} and μ_{j+1} are adjacent.

(a) If $n = 2$ we reach an immediate contradiction because by Lemma 3.2.1 μ_0 and μ_2 are not adjacent.

(b) If $n > 2$, and $j = 1$, then by Lemma 3.2.1, $\mu_0 \leftarrow \mu_2$ is in $\mathcal{H}$ and this edge is real. But then $\langle \mu_0, \mu_1, \mu_2, \mu_3 \rangle$ forms an inducing path and by Theorem 2.6.1 μ_3 is adjacent to μ_0. Note that μ_3 is not an endpoint because μ_0 is not adjacent to μ_n. Let μ_k be the first vertex after μ_1 that is not adjacent to μ_0; such a vertex is guaranteed to exist. Then $\langle \mu_0, \mu_1, \ldots, \mu_k \rangle$ forms an inducing path in $\langle$ and by Theorem 2.6.1 μ_k is adjacent to μ_0 which is a contradiction. Therefore, $j \neq 1$. A symmetric argument holds for the case $j = n - 1$.

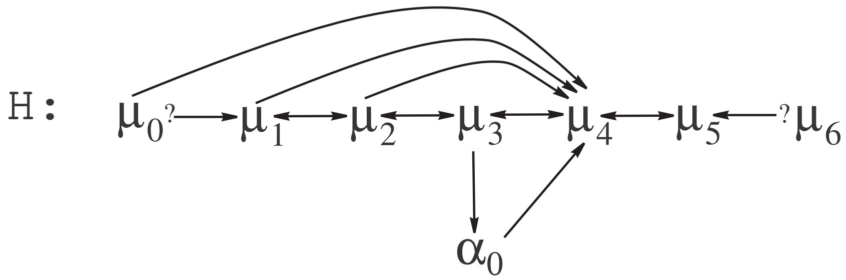


Figure 3.18: Diagram for Proof of Lemma 3.5.1. See text for details.

(c) If $n > 2, i < j < n-1$ then by Lemma 3.1.4 the $\{\mu_{j-1}, \mu_{j+1}\}$ edge is directed and real. Without loss of generality, suppose $\mu_{j-1} \to \mu_{j+1}$ occurs in $\mathcal{H}$ (see Figure 3.18). Then

$\langle \mu_{j-2}, \mu_{j-1}, \mu_j, \mu_{j+1} \rangle$ forms an inducing path in $\mathcal{H}$ and by Theorem 2.6.1 μ_{j-2} is adjacent to μ_{j+1}. Let μ_k be the closest vertex before μ_j that is not adjacent to μ_{j+1} or μ_0. If $\mu_k = \mu_0$ then we reach a contradiction: by Lemma 3.2.1 $\mu_0 \leftarrow \mu_{j+1}$ is in $\mathcal{H}$ and this edge is real; consequently, the non-ancestral configuration $\mu_0? \rightarrow \mu_1 \rightarrow \mu_{j+1} \rightarrow \mu_0$ occurs in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ (because the $\mu_1 \rightarrow \mu_{j+1} \rightarrow \mu_0$ edges are real). $\langle \mu_k, \mu_{k+1}, \ldots, \mu_j, \mu_{j+1} \rangle$ forms an inducing path and by Theorem 2.6.1 μ_k is adjacent to μ_{j+1} which is a contradiction.

A symmetric argument can be made for the case in which $\mu_j \leftarrow \alpha_0 \leftarrow \mu_{j+1}$ in $\mathcal{H}$. So, the configurations $\mu_j \rightarrow \alpha_0 \rightarrow \mu_{j+1}$ and $\mu_j \leftarrow \alpha_0 \leftarrow \mu_{j+1}$ cannot occur on a collider sub-path of a minimal j-connecting path in $\mathcal{H}$. $\square$

3.5.1 Examining the $\mu_i \rightarrow \alpha_0$ edge in $\mathcal{H}$

Lemma 3.5.2 *Let π be a minimal j-connecting path in graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Let $\mu = \langle \mu_0, \mu_1, \mu_2 \rangle$ be a collider sub-path of π such that neither μ_0 nor μ_2 are colliders on π. Suppose μ_1 is not in Z, and let α_0 be the first descendant of μ_1 on the ancestral path from μ_1 to Z in $\mathcal{H}$. Then the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real.*

Proof:

Suppose for a contradiction that the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is not real. Then there is at least one ancestral graph, $\mathcal{G}$, that has an arrowhead at μ_1 on the $\mu_1 \rightarrow \alpha_0$ edge. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_1 \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_0 and μ_2 are each adjacent to α_0.

There is at least one ancestral $\mathcal{G}$ that gave rise to $\mathcal{H}$ such that $\mu_1 \rightarrow \alpha_0$. In this $\mathcal{G}$, where μ_1 is a parent of α_0, μ_0 and μ_2 are either parents or spouses of α_0 because otherwise either $\mu_0? \rightarrow \mu_1 \rightarrow \alpha_0 \rightarrow \mu_0$ or $\mu_2 \leftarrow \alpha_0 \leftarrow \mu_1 \leftarrow ?\mu_2$ would form a non-ancestral configuration (see Figure 3.19). In other words, whenever $\mu_1 \rightarrow \alpha_0$ occurs in some $\mathcal{G}$, $\mu_0? \rightarrow \alpha_0 \leftarrow ?\mu_2$ also occurs. Note that $\mu_0? \rightarrow \alpha_0 \leftarrow ?\mu_2$ cannot occur in $\mathcal{H}$ because if it did, then the minimality of the path μ would be violated. Since α_0 is sometimes a collider and sometimes a non-collider on the path $\langle \mu_0, \alpha_0, \mu_2 \rangle$ in the ancestral graphs that gave rise to $\mathcal{H}$, μ_0 and μ_2 are

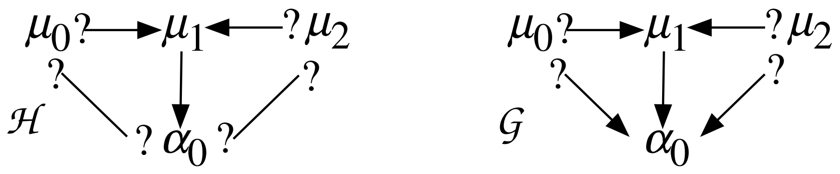


Figure 3.19: If $n = 2$ and the path μ is as outlined in Lemma 3.5.2, then μ_0 and μ_2 are adjacent to α_0.

adjacent. But this is a contradiction because by Lemma 3.2.1 μ_0 and μ_n are not adjacent. Therefore, the $\mu_1 \to \alpha_0$ edge is real. □

Corollary 3.5.1 *(to Lemma 3.1.4) Let μ_i be a collider on a minimal j-connecting path in some joined graph $\mathcal{H}$, and α_0 be the first descendant of μ_i on the ancestral path from μ_i to a vertex in Z, as shown in Figure 3.20. If α_0 is another collider on the path μ, then the $\mu_i \to \alpha_0$ edge is real.*

Proof:

This result follows directly from Lemma 3.1.4 because by Lemma 3.1.4 if μ_i is a parent to any other vertex on the collider sub-path, this edge is real. □

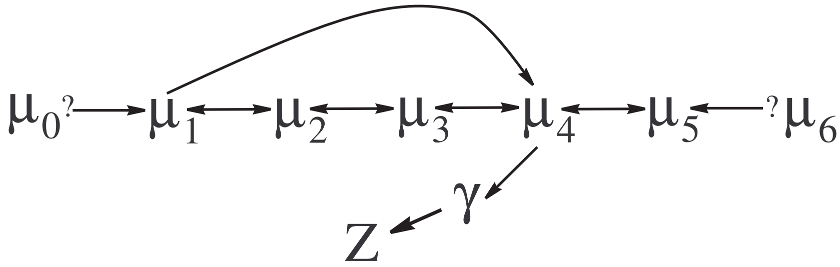


Figure 3.20: Diagram for Proof of Corollary 3.5.1. See text for details.

Lemma 3.5.3 *Let π be a minimal j-connecting path in the graph $\mathcal{H}$, formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Let $\mu = \langle \mu_0, \ldots, \mu_n \rangle$ be a collider sub-path of π such that neither μ_0 nor μ_n are colliders on π. Suppose μ_1 is not in Z, and let α_0 be the first descendant of μ_1 on the ancestral path from μ_1 to Z in $\mathcal{H}$ with $\mu_0 ? \!-\! \alpha_0$. Then the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is real.*

Proof:

Suppose for a contradiction that the $\mu_1 \rightarrow \alpha_0$ edge in $\mathcal{H}$ is not real. Then there is at least one ancestral graph, $\mathcal{G}$, that has an arrowhead at μ_1 on the $\mu_1 \rightarrow \alpha_0$ edge. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_1 \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_0 and μ_2 are each adjacent to α_0.

Case I: $n = 2$, i.e. $\mu_0 ? \rightarrow \mu_1 \leftarrow ? \mu_2$ is in $\mathcal{H}$.

By Lemma 3.5.2 the edge $\mu_1 \rightarrow \alpha_0$ in $\mathcal{H}$ is real which is a contradiction.

Case II: $n \geq 3$.

By Corollary 3.5.1, $\alpha_0 \neq \mu_t, t \neq i$. I will show that for each $\mu_r, 2 \leq r \leq (n-1)$:

1. $\alpha_0 ? \!-\! \mu_r$ in $\mathcal{H}$.
2. $\mu_0 \leftarrow \mu_r$ in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and the edge is real (by Lemma 2.5.2).

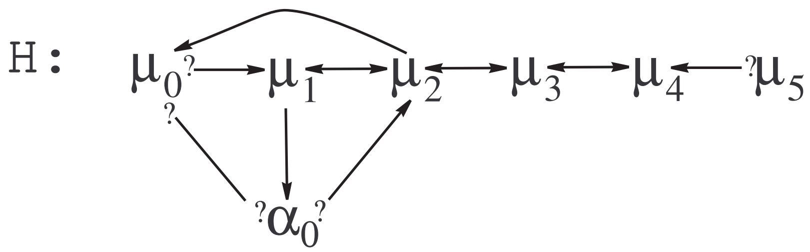


Figure 3.21: Diagram (a) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

Inductive Case: $2 \leq r < n$ Let μ_r be the first vertex after μ_1 such that the following two conditions do *not* hold:

1. $\alpha_0 ?\!-\!\!-\, \mu_r$ in $\mathcal{H}$.
2. $\mu_0 \leftarrow \mu_r$ in $\mathcal{H}$ and this edge is real.

If μ_r is not adjacent to α_0 and $r = 2$ then $\langle \mu_0, \mu_1, \alpha_0 \rangle$ forms an unshielded collider. If μ_r is not adjacent to α_0 and $r > 2$, then $\langle \mu_r, \mu_{r-1}, \ldots, \mu_2, \mu_1, \alpha_0 \rangle$ forms a discriminating path for μ_1 and this path is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, which is a contradiction since by Lemma 2.1.1 $\mu_1 \rightarrow \alpha_0$ would be real. Therefore μ_r is adjacent to α_0.

I will now show that $\alpha_0 ?\!-\!\!-\, \mu_r$ is in $\mathcal{H}$. Suppose for a contradiction that $\alpha_0 ?\!\rightarrow \mu_r$ is in $\mathcal{H}$.

(a) For $r < n$ if $\alpha_0 \rightarrow \mu_r$ is in $\mathcal{H}$, then $\langle \mu_1, \mu_2, \ldots, \mu_r \rangle$ forms an inducing path in $\mathcal{H}$ (because each μ_i, $2 \leq i < r$ is a parent of μ_0, and $\mu_1 \rightarrow \alpha_0 \rightarrow \mu_r$ is in $\mathcal{H}$) and by Theorem 2.6.1 μ_r is adjacent to μ_0, and by Corollary 2.5.2 $\mu_0 \leftarrow \mu_r$ is in $\mathcal{H}$ and this edge is real. See Figure 3.22.

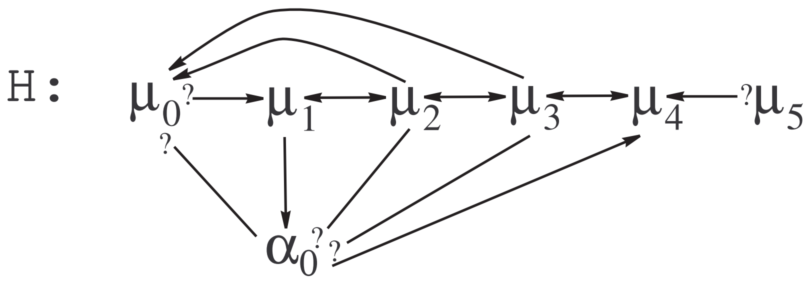


Figure 3.22: Diagram (b) for Proof of Lemma 3.5.3 with $r = n = 5$. See text for details.

Consider the case where $r = n$. If $\alpha_0 \rightarrow \mu_r$ is in $\mathcal{H}$ then since μ_0 and μ_r are not adjacent, $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider that is present in all $\mathcal{G}$ joined to form $\mathcal{H}$. By the definition of μ_r each of $\{\mu_2, \ldots, \mu_{n-1}\}$ is a parent of μ_0 in every $\mathcal{G}$ joined to form $\mathcal{H}$. Consider the maximal ancestral graph $\mathcal{G}^*$, used to form $\mathcal{H}$, in which $\mu_1 \rightarrow \alpha_0$ (See Figure

3.23); such a $\mathcal{G}$ exists because $\mu_1 \to \alpha_0$ in $\mathcal{H}$. In $\mathcal{G}^*$, $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider so α_0 is a parent of at least one of μ_0 or μ_n. In fact, by Lemma 2.5.3 $\mu_0 ? \to \alpha_0$ occurs in $\mathcal{G}^*$ so $\alpha_0 \to \mu_n$ occurs in $\mathcal{G}^*$. But then μ forms an inducing path in $\mathcal{G}^*$ and by the maximality of $\mathcal{G}^*$, μ_0 is adjacent to μ_n. Consequently μ_0 is adjacent to μ_n in $\mathcal{H}$, which is a contradiction. Therefore if $\alpha_0 \to \mu_r$ is in $\mathcal{H}$, then $r < n$.

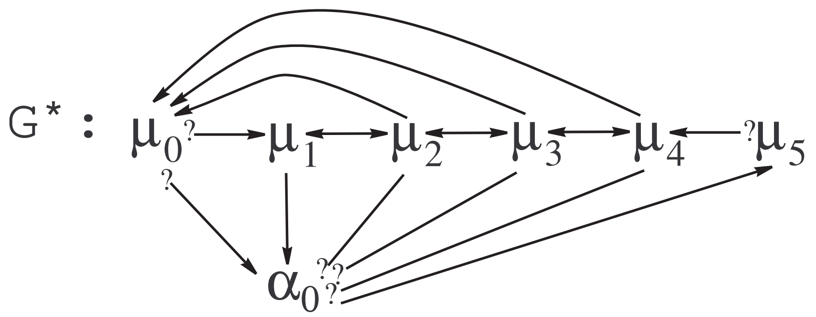


Figure 3.23: Diagram (c) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

(b) If $\alpha_0 \leftrightarrow \mu_r$ is in $\mathcal{H}$, then $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a non-collider in $\mathcal{H}$ (with an arrowhead at α_0 on the $\{\alpha_0, \mu_r\}$ edge in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$). We know that there is at least one $\mathcal{G}$ in which $\mu_0 ? \to \alpha_0$ occurs and in this graph $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a collider. We also know that there is at least one $\mathcal{G}^{*}*$ in which $\mu_0 \leftarrow \alpha_0$ and hence $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a non-collider in $\mathcal{G}^{*}*$. Consequently, μ_0 is adjacent to μ_r (see Figure 3.24). If $r = n$ then we reach a contradiction because by assumption μ_0 and μ_n are not adjacent. Thus, if $\alpha_0 \leftrightarrow \mu_r$ then $r < n$ and by Corollary 2.5.2, $\mu_0 \leftarrow \mu_r$ and this edge is real.

So, if $\alpha_0 ? \to \mu_r$ is in $\mathcal{H}$, then $r < n$, $\mu_0 \leftarrow \mu_r$ and this edge is real. But then $\langle \mu_{r+1}, \mu_r, \alpha_0, \mu_0 \rangle$ forms a discriminating path for α_0 in $\mathcal{H}$, and discriminates $\langle \mu_0, \alpha_0, \mu_r \rangle$ to be a non-collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. However, this is a contradiction for the following reason:

Consider a $\mathcal{G}^*$ in which $\mu_1 \to \alpha_0$. By Lemma 2.5.3 (triangle rule), $\mu_0 ? \to \alpha_0$ is in $\mathcal{G}^*$. Because the $\mu_0 \leftarrow \mu_r$ edge is real, by Lemma 2.5.3 there is also an arrowhead at α_0

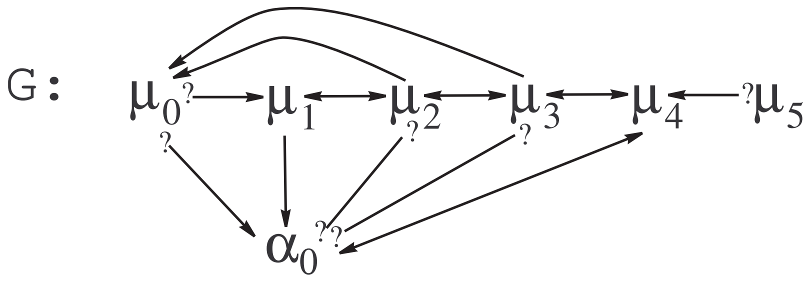


Figure 3.24: Diagram (e) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

on the $\{\alpha_0, \mu_r\}$ edge, i.e. $\alpha_0 \leftrightarrow \mu_r$. But then $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a collider in $\mathcal{G}^*$, which is a contradiction. See Figure 3.25

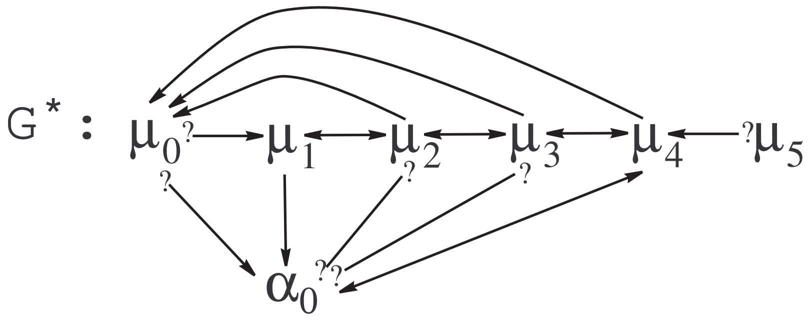


Figure 3.25: Diagram (e) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

Therefore, I have shown that $\alpha_0 ? \rightarrow \mu_r$ is not in $\mathcal{H}$. Consequently we can infer that $\alpha_0 ? - \mu_r$ is in $\mathcal{H}$. I will now show that μ_r is adjacent to μ_0 (which leads to an immediate contradiction because of the definition of μ_r).

Currently we know the following about $\mathcal{H}$ (see Figure 3.26):

a) $\mu_0 ? - \alpha_0$,

b) $\mu_0 \leftarrow \mu_i$ for $2 \leq i < r$ and these edges are real, and

c) $\alpha_0 ? - \mu_i$ for $2 \leq i \leq r$

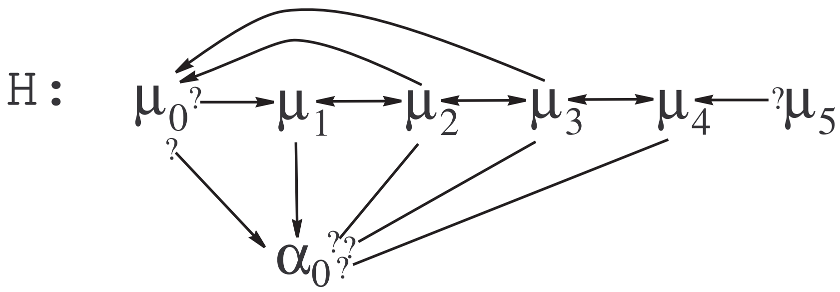


Figure 3.26: Diagram (f) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

Suppose for a contradiction that μ_r is not adjacent to μ_0. Then $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms an unshielded non-collider in $\mathcal{H}$ and this non-collider is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

Consider the $\mathcal{G}$ in which $\mu_1 \to \alpha_0$ occurs. In this $\mathcal{G}$, $\mu_0 ? \to \alpha_0$ also occurs, so $\mu_i \to \mu_0 ? \to \alpha_0 ? - ?\mu_i$ occurs in $\mathcal{H}$ for $2 \leq i < r$. By a simple triangle argument, and the fact that $\mathcal{G}$ is ancestral, we can infer an arrowhead at α_0 on the $\{\mu_i, \alpha_0\}$ edge for every i, $2 \leq i < r$. I have already shown that $\langle \mu_0, \alpha_0, \mu_r \rangle$ is a non-collider in all graphs that gave rise to $\mathcal{H}$. Since $\mu_0 ? \to \alpha_0$ is in $\mathcal{G}$ we can infer that $\alpha_0 \to \mu_r$ is in $\mathcal{G}$ (else $\langle \mu_0, \alpha_0, \mu_r \rangle$ forms a collider). See Figure 3.27.

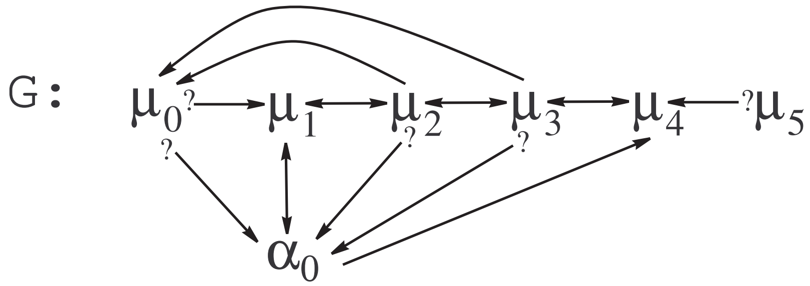


Figure 3.27: Diagram (g) for Proof of Lemma 3.5.3 with $r = 4$. See text for details.

If there is no arrowhead at μ_1 on the $\{\mu_1, \alpha_0\}$ edge in $\mathcal{G}$, i.e. $\mu_1 \to \alpha_0$, then $\pi(\mu_0, \mu_r)$ forms an inducing path and by the maximality of $\mathcal{G}$, μ_r would be adjacent to μ_0 which would be a contradiction. Therefore $\mu_1 \leftrightarrow \alpha_0$ occurs in this $\mathcal{G}$ which is a contradiction. □

The following proof is the same as the proof of Lemma 2.6.4. The only difference between the proofs is that Lemma 2.6.4 deals with minimal inducing paths (in which each non-endpoint is an ancestor of an endpoint), and Lemma 3.5.4 deals with minimal j-connecting paths. Since the proof of Lemma 2.6.4 does not use the fact that the path is inducing, these proofs are identical. The proof is included here for completeness.

Lemma 3.5.4 *Let π be a minimal j-connecting path in graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Let $\mu = \langle \mu_0, \ldots, \mu_2 \rangle$ be a collider sub-path of π such that μ_0 and μ_n such that neither vertex is a collider on π. Suppose μ_1 is not in Z, and let α_0 be the first descendant of μ_1 on the ancestral path from μ_1 to Z in $\mathcal{H}$ with $\mu_0 ? \to \alpha_0$. Then the $\mu_1 \to \alpha_0$ edge in $\mathcal{H}$ is real.*

Proof:

Suppose for a contradiction that the $\mu_1 \to \alpha_0$ edge is not real. Then there is at least one ancestral graph $\mathcal{G}$ in which the $\{\mu_1, \alpha_0\}$ edge is oriented $\mu_1 \leftrightarrow \alpha_0$. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_1 \to \alpha_0$ edge is shielded from both sides, i.e. μ_0 and μ_2 are each adjacent to α_0.

Case I: $n = 2$, i.e. $\mu_0 ? \to \mu_1 \leftarrow ? \mu_2$ is in $\mathcal{H}$. By Lemma 3.5.2 the $\mu_1 \to \alpha_0$ edge in $\mathcal{H}$ is real which is a contradiction.

Case II: $n \geq 3$. By Corollary 3.5.1, α_0 is not equal to another $\mu_t, t \neq i$. Since μ is minimal, $\alpha_0 \leftrightarrow \mu_j$, $2 \leq j \leq n$ does not occur in $\mathcal{H}$ because in this case $\langle \mu_0, \alpha_0, \mu_j, \ldots, \mu_n \rangle$ would violate the minimality of μ. Let μ_k be the vertex furthest from μ_1 such that μ_i is adjacent to α_0 for $0 \leq i \leq k$. Since μ_2 is adjacent to α_0, $r \geq 2$.

Claim: $\alpha_0 \leftarrow \mu_i$ is in $\mathcal{H}$ for $1 \leq i \leq k$.

Suppose that there exists an $i \leq k$ such that $\alpha_0 - ?\mu_k$ is in $\mathcal{H}$. Let μ_r be the vertex closest to μ_1 such that $\alpha_0 - ?\mu_r$ is in $\mathcal{H}$. By the definition of μ_r, $\alpha_0 \leftarrow \mu_{r-1}$ is in $\mathcal{H}$.

If $r = 2$, then by assumption $\alpha_0 \leftarrow \mu_1$ and the claim is true. There are now two cases to consider:

(i) μ_r is adjacent to μ_0 (See Figure 3.28).

If μ_r is adjacent to μ_0, then by Corollary 2.5.2 $\mu_0 \leftarrow \mu_r$, and this edge is real. Hence by Lemma 2.5.3, since $\mu_0 ?\!\rightarrow \alpha_0 -\!? \mu_r \rightarrow \mu_0$ it follows that there is an arrowhead at α_0 on the $\{\mu_r, \alpha_0\}$ edge in $\mathcal{H}$ (i.e. $\alpha_0 \leftarrow\!? \mu_r$), which is a contradiction. Therefore, μ_r is not adjacent to μ_0.

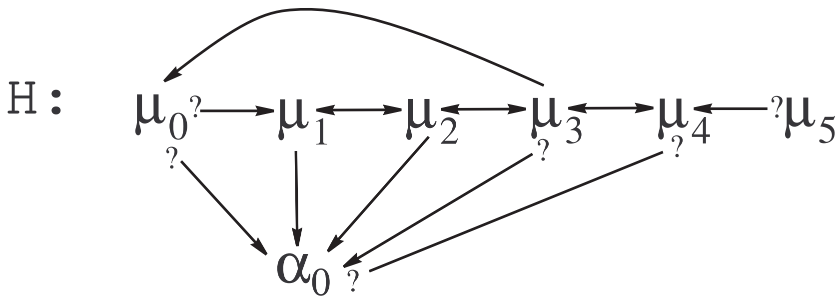


Figure 3.28: Diagram (a) for Proof of Lemma 3.5.4 with $r = 4$. See text for details.

(ii) μ_r is not adjacent to μ_0 (See Figure 3.29).

In this case, $\mu_0 ?\!\rightarrow \alpha_0 -\!? \mu_r$ forms an unshielded non-collider that is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Hence $\alpha_0 \rightarrow \mu_r$ is in $\mathcal{H}$ and because by definition of the join operation $\mu_0 ?\!\rightarrow \alpha_0$ occurs in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$, the $\alpha_0 \rightarrow \mu_r$ edge in $\mathcal{H}$ is real. Since $\alpha_0 \leftarrow \mu_{r-1}$ is in $\mathcal{H}$ (by the definition of μ_r), there exists at least one maximal ancestral graph $\mathcal{G}^*$ in which $\alpha_0 \leftarrow \mu_{r-1}$. In this graph $\mathcal{G}^*$, the triangle $\mu_r ?\!\rightarrow \mu_{r-1} \rightarrow \alpha_0 \rightarrow \mu_r$ occurs, which is a contradiction because $\mathcal{G}^*$ is ancestral.

Consequently for $1 \leq i \leq k$, $\mu_k \rightarrow \alpha_0$ occurs in $\mathcal{H}$. If $k < n$, then μ_{k+1} exists and is not adjacent to α_0. However, $\langle \mu_0, \mu_{k+1}, \ldots, \mu_2, \mu_1, \alpha_0 \rangle$ forms a discriminating path for μ_1 in $\mathcal{H}$ and by Lemma 2.5.4, this path is discriminating in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$. Hence $k = n$, which is a contradiction because $\mu_0 ?\!\rightarrow \alpha_0 \leftarrow\!? \mu_n$ violates the minimality of the path μ. Consequently, the $\mu_1 \rightarrow \alpha_0$ edge is real.

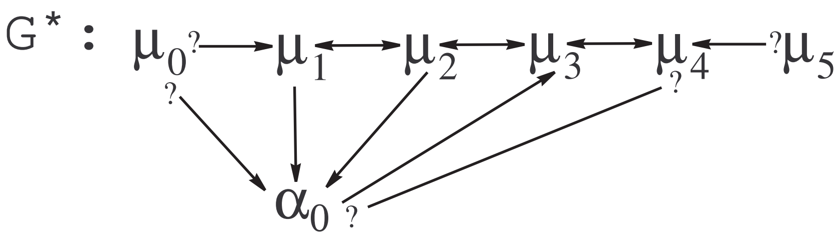


Figure 3.29: Diagram (b) for Proof of Lemma 3.5.4 with $r = 4$. See text for details.

□

Lemma 3.5.5 *Let π be a minimal j-connecting path in graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$. Let $\mu = \langle \mu_0, \ldots, \mu_n \rangle$ be a collider sub-path of π such that neither μ_0 nor μ_n is a collider on π. Suppose μ_i is not in Z, and let α_0 be the first descendant of μ_i on the ancestral path from μ_i to Z in $\mathcal{H}$. Then the $\mu_i \rightarrow \alpha_0$ edge is real.*

Proof:

For a contradiction, suppose the $\mu_i \rightarrow \alpha_0$ edge is not real. Then there is at least one ancestral graph $\mathcal{G}$ that has an arrowhead at μ_i on the $\mu_i \rightarrow \alpha_0$ edge. Since all $\mathcal{G}$s that gave rise to $\mathcal{H}$ are Markov equivalent, the $\mu_i \rightarrow \alpha_0$ edge is shielded from both sides, i.e. μ_{i-1} and μ_{i+1} are also adjacent to α_0.

Case I: $n = 2$. By Lemma 3.5.2 if $n = 2$, then the $\mu_1 \rightarrow \alpha_0$ edge is real.

Case II: $n \geq 3$. For $n \geq 3$, Lemmas 3.5.3 and 3.5.4 show that if $i = 1, (n-1)$, then the $\mu_i \rightarrow \alpha_0$ edge is real. Now I consider the case in which $n \geq 3, 1 > i > (n-1)$. Note that by Corollary 3.5.1 $\alpha_0 \neq \mu_t, t \neq i$. If $\mu_j, 0 \leq j < i$ is a spouse of α_0, and $\mu_k, i < k \leq n$ is a spouse of α_0 then the path $\langle \mu_0, \mu_1, \ldots, \mu_j, \alpha_0, \mu_k, \mu_{k+1}, \ldots, \mu_n \rangle$ is more minimal than μ. Without loss of generality assume that $\mu_j \leftrightarrow \alpha_0, 0 \leq j < i$ does not occur in $\mathcal{H}$ and that μ_0 is not a parent of α_0.

By Lemma 3.5.1, neither μ_{i-1} nor μ_{i+1} are children of α_0. Therefore, $\mu_{i-1} -?\alpha_0$ is in $\mathcal{H}$. Unless μ_{i-2} is adjacent to α_0, $\langle \mu_{i-2}, \mu_{i-1}, \mu_i, \alpha_0 \rangle$ forms a discriminating path for μ_i (because by Lemma 2.1.1 $\langle \mu_{i-2}, \mu_{i-1}, \mu_i \rangle$ forms an unshielded non-collider which would imply that the tail on the $\{\mu_{i-1}, \alpha_0\}$ edge is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ and hence the $\mu_{i-1} \to \alpha_0$) edge is real). Thus, μ_{i-2} is adjacent to α_0.

By assumption $\mu_{i-2} \leftrightarrow \alpha_0$ does not occur in $\mathcal{H}$, so either $\mu_{i-1} \to \alpha_0$ or $\mu_{i-2} ?\!-\!\!- \alpha_0$ is in $\mathcal{H}$. Suppose there is a tail at α_0 on the $\{\mu_{i-2}, \alpha_0\}$ edge. If the tail is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ then $\mu_{i-2} \leftarrow \alpha_0$ occurs in $\mathcal{H}$ and this edge is real. But then by Lemma 2.5.2 $\mu_{i-1} \leftarrow ?\alpha_0$ is in $\mathcal{H}$. Since $\mu_{i-1} \leftrightarrow \alpha_0$ is excluded by assumption, $\mu_{i-1} \leftarrow \alpha_0$ is in $\mathcal{H}$ which is a contradiction by Lemma 3.5.1. If the tail at α_0 on the $\{\mu_{i-2}, \alpha_0\}$ edge is not present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ then μ_{i-2} is adjacent to μ_i to shield $\langle \mu_{i-2}, \alpha_0, \mu_i \rangle$. Furthermore, since μ_i is not in Z, by Lemma 3.1.4 $\mu_{i-2} \to \mu_i$ and this edge is real. But then by Lemma 2.5.2 $\mu_{i-2} \to \alpha_0$, which is a contradiction. Therefore, $\mu_{i-2} ?\!-\!\!- \alpha_0$ is not in $\mathcal{H}$, and consequently, $\mu_{i-2} \to \alpha_0$ is in $\mathcal{H}$.

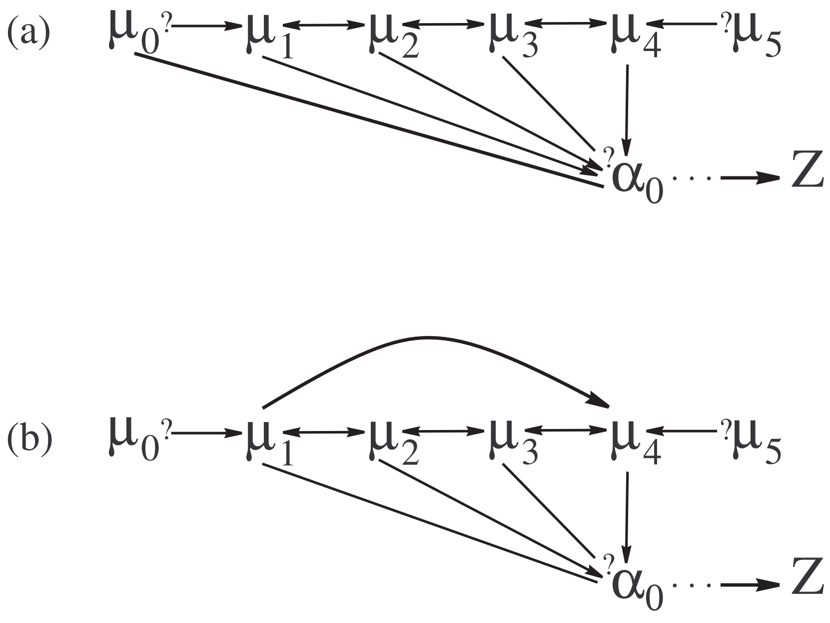


Figure 3.30: Diagram for Proof of Lemma 3.5.5. See text for details.

Let μ_k be the vertex closest to μ_0 that is before μ_i such that μ_k is not a parent of α_0. Consider the case where $\mu_k = \mu_0$ (see Figure 3.30(a)): Then μ_1 is a parent of α_0

in $\mathcal{H}$. If $\mu_0 \leftarrow \alpha_0$ is in $\mathcal{H}$, then $\langle \mu_0, \mu_1, \mu_2 \rangle$ forms an inducing path and by Theorem 2.6.1 μ_0 is adjacent to μ_2. But then $\langle \mu_0, \mu_1, \mu_2, \mu_3 \rangle$ forms an inducing path in $\mathcal{H}$ and by Theorem 2.6.1 μ_0 is adjacent to μ_3. By repeated applications of this inducing path argument, $\langle \mu_0, \ldots, \mu_n \rangle$ forms an inducing path in $\mathcal{H}$ and by Theorem 2.6.1 μ_0 is adjacent to μ_n which is a contradiction. Thus, $\mu_0 \leftarrow \alpha_0$ is not in $\mathcal{H}$; and furthermore, by assumption $\mu_0 \leftrightarrow \alpha_0$ is not in $\mathcal{H}$.

Therefore $\mu_0 -?$ is in $\mathcal{H}$. By the definition of μ_k, $\mu_0 \rightarrow \alpha_0$ is ruled out because $\mu_k = \mu_0$ is not a parent of α_0, so $\mu_0 - \alpha_0$ is in $\mathcal{H}$. By Lemma 2.5.1 the $\{\mu_0, \alpha_0\}$ edge is not real, so μ_0 is adjacent to μ_i. But this is a contradiction because by Lemma 3.2.1 μ_0 and μ_i are not adjacent (since μ_0 is not a collider, and μ_i is not in Z). Therefore, $0 < k < i$.

If μ_k is not adjacent to α_0 then $\langle \mu_k, \mu_{k+1}, \ldots, \mu_{i-1}, \alpha_0 \rangle$ discriminates $\langle \mu_{i-2}, \mu_{i-1}, \alpha_0 \rangle$ to be a non-collider in $\mathcal{H}$ and $\langle \mu_k, \mu_{k+1}, \ldots, \mu_{i-1}, \mu_i, \alpha_0 \rangle$ forms a discriminating path for μ_i, which is a contradiction (see Figure 3.30(a)). Thus, μ_k is adjacent to α_0. By assumption $\mu_k \leftrightarrow \alpha_0$ is ruled out; by Lemma 3.5.1 $\mu_k \leftarrow \alpha_0$ is ruled out; and by the definition of μ_k, $\mu_k \rightarrow \alpha_0$ is ruled out. Therefore $\mu_k - \alpha_0$ is in $\mathcal{H}$. Since an arrowhead meets the undirected edge $\{\mu_k, \alpha_0\}$, by Lemma 2.5.1 μ_k is adjacent to μ_i. Furthermore, since μ_i is not in Z, by Lemma 3.1.4 $\mu_k \rightarrow \mu_i$ and this edge is real. But now we reach a contradiction: because the configuration $\mu_k \rightarrow \mu_i \rightarrow \alpha_0 - \mu_k$ is in $\mathcal{H}$, by Lemma 2.5.2, there is an arrowhead at α_0 on the $\{\mu_k, \alpha_0\}$ edge.

Since any joined graph $\mathcal{H}$, for which the edge $\mu_i \rightarrow \alpha_0$ is *not* real leads to a contradiction, the edge $\mu_i \rightarrow \alpha_0$ is real. $\Box$

Lemma 3.5.6 *Let $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent maximal ancestral graphs $\mathcal{G}$ and let π be a minimal j-connecting path between α and β given Z in $\mathcal{H}$. If μ is a collider sub-path of π with endpoints μ_0 and μ_n such that neither μ_0 nor μ_n are colliders along π, then μ_i, $0 < i < n$ is an ancestor of Z in all $\mathcal{G}$ joined to form $\mathcal{H}$.*

Proof:

Since μ is a collider path that is j-connecting given Z in $\mathcal{H}$, every non-endpoint vertex μ_i is an ancestor of Z in $\mathcal{H}$. Furthermore, by the definition of the join operation, the path μ is a collider path in all ancestral graphs $\mathcal{G}$ that gave rise to $\mathcal{H}$.

I will show that all the edges on the ancestral path from $\mu_i, 0 < i < n$ to Z are real. Hence, I will have shown that the path μ in $\mathcal{H}$ would be j-connecting given Z in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

To show: All edges on the ancestral path from μ_i to Z, i.e. $\langle \mu_i, \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m \rangle$ where α_m is in Z, are real.

Lemma 3.5.5 states that α_0, is a child of μ_i in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ (i.e. $\mu_i \rightarrow \alpha_0$ is real). I will use an inductive proof to show that all the subsequent edges on the ancestral path from μ_i to Z are also real.

Suppose that $\alpha_0 \rightarrow \alpha_1$ is not real (see Figure 3.31). Then μ_i is adjacent to α_1 since otherwise $\langle \mu_i, \alpha_0, \alpha_1 \rangle$ is an unshielded non-collider and consequently the $\alpha_0 \rightarrow \alpha_1$ edge is real. By Lemma 2.5.3 $\mu_i \rightarrow \alpha_1$ is in $\mathcal{H}$, but then the minimality of the path is violated since we can take a directed path from μ_i to Z that bypasses α_0. So, the $\alpha_0 \rightarrow \alpha_1$ is real.

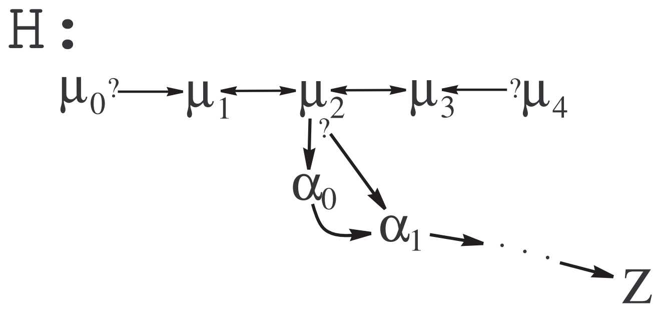


Figure 3.31: If the path μ is as outlined in Lemma 3.5.2, then μ_2 and α_1 are adjacent, and μ_2 is a parent of α_1.

Assume that for $0 < k < m$, the $\alpha_{k-1} \rightarrow \alpha_k$ edge is real and consider the $\alpha_k \rightarrow \alpha_{k+1}$ edge. If the edge $\alpha_k \rightarrow \alpha_{k+1}$ in $\mathcal{H}$ is not real, then α_{k-1} is adjacent to α_{k+1} since otherwise $\langle \alpha_{k-1}, \alpha_k, \alpha_{k+1} \rangle$ is an unshielded collider. By Lemma 2.5.3 $\alpha_{k-1} \rightarrow \alpha_{k+1}$ is in $\mathcal{H}$, but then the minimality of the path is violated since we can take the path from μ_i to Z that bypasses α_k. Consequently, the $\alpha_k \rightarrow \alpha_{k+1}$ edge is real. So, by induction, the edges between $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m$ are real and form an ancestral path from α_0 to Z.

It has been shown that if the path μ is j-connecting in $\mathcal{H}$ given the vertex set Z, then it is also j-connecting given Z in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. $\square$

3.6 From j-connection to m-connection

In this chapter I show that j-connection in a joined graph implies m-connection in all the maximal ancestral graphs that gave rise to the joined graph. Before I present the proof of this result, I present results regarding Markov equivalence of ancestral graphs that will facilitate the proof that m-connection in $\mathcal{G}$ implies j-connection in $\mathcal{H}$.

Corollary 3.6.1 *(to Lemma 3.2.1) Let π be a minimal m-connecting path between α and β given Z in the ancestral graph $\mathcal{G}$, and δ is either an endpoint or a non-collider consisting of directed edges along π. If δ is adjacent to γ, a non-consecutive vertex along π, then $\delta \leftarrow \gamma$ is in $\mathcal{G}$, γ is a collider on π and γ is in Z.*

Proof:
Ancestral graphs are a subset of joined graphs so this lemma is true by Lemma 3.2.1 after observing that arrowheads do not meet undirected edges in ancestral graphs. $\square$

3.7 Discriminating Paths in Joined Graphs

Discriminating paths can identify which shielded vertices in a graph are important in determining Markov equivalence of ancestral graphs. However, not all paths that are discriminating in one graph are discriminating in another graph. Consider the example given in Figure 3.32(i).

Note that q is a collider on the path $\langle x, q, B, y\rangle$ in $\mathcal{H}_1$, but not in $\mathcal{H}_2$ so $\langle x, q, B, y\rangle$ forms a discriminating path in $\mathcal{H}_1$, but not in $\mathcal{H}_2$. If we knew $\mathcal{H}_2$, then we would not know whether B is a collider or a non-collider on the path j-connecting x and y in any graph Markov equivalent to $\mathcal{H}_2$. However, note that $\langle x, q, B, y\rangle$ is not a minimal j-connecting path between x and y given Z in $\mathcal{H}_2$ because $\langle x, B, y\rangle$ is a shorter j-connecting path between x and y given Z. Furthermore, note that B is an unshielded collider on the path $\langle x, B, y\rangle$

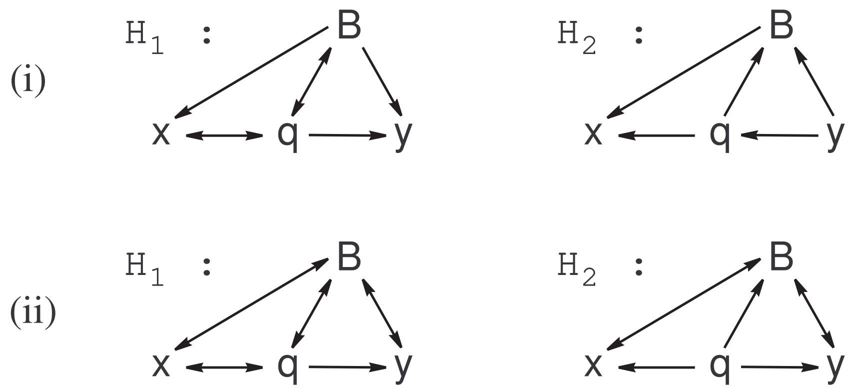


Figure 3.32: Two examples of joined graphs that are Markov equivalent and $\langle x, q, B, y\rangle$ forms a discriminating path in $\mathcal{H}_1$, but not in $\mathcal{H}_2$. See text for further explanation.

in $\mathcal{H}_1$ and B is a non-collider on any j-connecting path between x and y on which B is a vertex, in any graph Markov equivalent to $\mathcal{H}_2$.

If a discriminating path $U = \langle x, q_1, \ldots, q_p, B, y\rangle$ is present in all graphs Markov equivalent to $\mathcal{G}$ then in all graphs $\mathcal{G}^*$ Markov equivalent to $\mathcal{G}$: (i) $\langle x, q_1, \ldots, q_p, B\rangle$ forms a collider path, and (ii) $\{q_1, \ldots, q_p\}$ are all parents of y. If no vertex on U except B is shielded, then trivially $\langle x, q_1, \ldots, q_p, B\rangle$ forms a collider path and an inductive proof can be used to show that there is a path that discriminates each $q_i, 1 \leq i \leq p$ to be a parent of y. The following lemma and corollaries will be used to show that whenever U is part of a minimal m-connecting path given Z, even if some of the q_i's on U are shielded, there is a discriminating path that discriminates such a q_i to be a collider. I will use the notation $U(\alpha, \beta)$ to denote the sub-path of U formed by the vertices between α and β on U. Note that the following results are very similar to those presented in Richardson and Spirtes (2000).

Lemma 3.7.1 *In the joined graph $\mathcal{H}$, formed by joining sets of Markov equivalent maximal ancestral graphs, if U is a minimal j-connecting path between x and y given Z, then there is no pair of distinct vertices $\{B, J\}$ such that B is a shielded collider on the discriminating path $U(I, K)$ for J and J is a shielded collider on the discriminating path $U(A, C)$ for B.*

Proof:

For a contradiction, suppose B is a shielded collider on $U(I,K)$ (B is a collider because it is on $U(I,K)$ but is not equal to J) and J is a shielded collider on $U(A,C)$. Since B and J are distinct, $B \neq J$. Furthermore, $C \neq K$: otherwise either B would not lie on $U(I,K)$ or J would not lie on $U(A,C)$ because $B \neq J$ (see Figure 3.33).

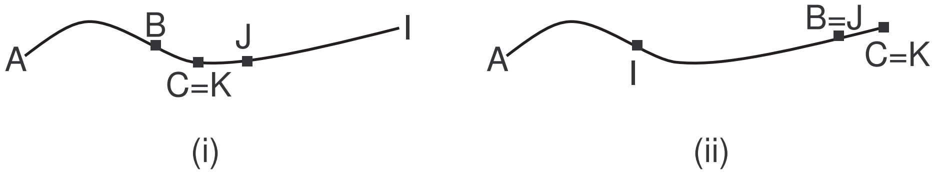


Figure 3.33: Diagram 1 for proof of Lemma 3.7.1. See text for further explanation.

$J \neq A, C$ because J is a collider on $U(A,C)$, and similarly, $B \neq I, K$. Since J is a shielded collider on $U(A,C)$, J is a parent of C. Consequently, $C \neq I$ because otherwise $U(I,K)$ forms an inducing path and by Theorem 2.6.1 I is adjacent to K which is a contradiction (see Figure 3.34).

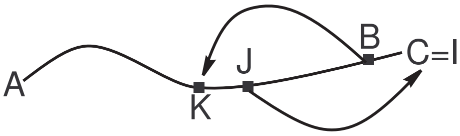


Figure 3.34: Diagram 2 for proof of Lemma 3.7.1. See text for further explanation.

Similarly, B is a parent of K, and $K \neq A$. Therefore, (i) K lies on $U(A,C)$ and (ii) C lies on $U(I,K)$, which is a contradiction: by (i) K is a parent of C, but by (ii) C is a parent of K (see Figure 3.35). $\square$

Corollary 3.7.1 *In the joined graph $\mathcal{H}$, formed by joining a set of Markov equivalent maximal ancestral graphs, if U is a minimal j-connecting path between A and B given Z, then*

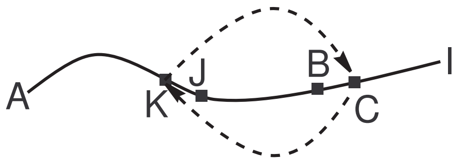


Figure 3.35: Diagram 3 for proof of Lemma 3.7.1. See text for further explanation.

there is no triple of distinct vertices $\{X, Y, Z\}$ on U such that X is a shielded collider on the discriminating path for Y, Y is a shielded collider on the discriminating path for Z on U, and Z is between X and Y on U.

Proof:

Suppose for a contradiction that such distinct vertices $\{X, Y, Z\}$ exist in $\mathcal{H}$ on the j-connecting path U as shown in Figure 3.36. Since Z is between X and Y, Z is a shielded collider on the discriminating path for Y and by assumption Y is a shielded collider on the discriminating path for Z. But this is a contradiction by Lemma 3.7.1. □

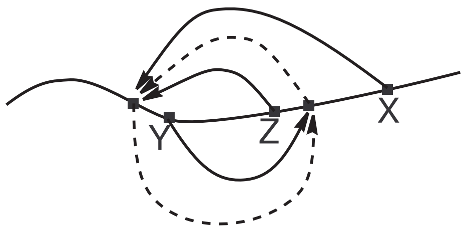


Figure 3.36: Diagram for proof of Corollary 3.7.1. See text for further explanation.

Corollary 3.7.2 *In the joined graph $\mathcal{H}$, formed by joining sets of Markov equivalent maximal ancestral graphs, if U is a j-connecting path between x and y given Z, then there is no quadruple of distinct vertices $\langle A_i, A_{r+1}, A_{i+1}, A_r \rangle$, in that order, on U such that A_i is a collider on the discriminating path for A_{i+1} on U; and A_r is a collider on the discriminating path A_{r+1}.*

Proof:

For a contradiction, suppose that such distinct vertices $\langle A_i, A_{r+1}, A_{i+1}, A_r \rangle$ exist on U in $\mathcal{H}$. Let $U(A_0, A_t)$ be the discriminating path for A_{i+1} and $U(A_n, A_s)$ be the discriminating path for A_{r+1} (see Figure 3.37). Note that possibly $A_r = A_t$ or $A_i = A_s$; likewise $A_r = A_n$ and $A_i = A_0$ are also possible.

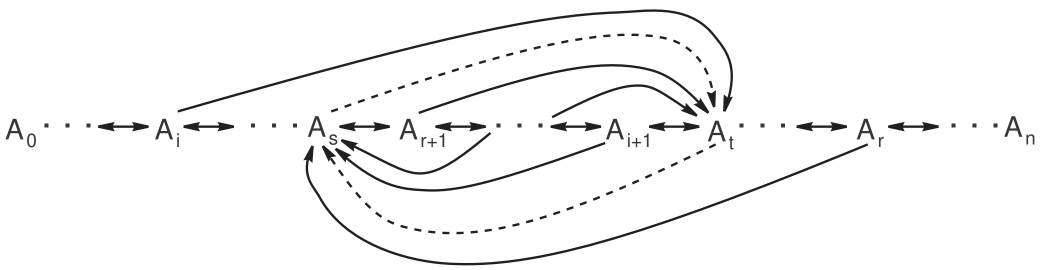


Figure 3.37: Diagram for proof of Corollary 3.7.2. See text for further explanation.

Then A_{i+1} is a shielded collider on $U(A_0, A_t)$ because A_{i+1} is between A_r and A_{r+1} which are on $U(A_0, A - t)$, and similarly A_{r+1} is a shielded collider on $U(A_n, A_s)$ which, by Lemma 3.7.1 is a contradiction. □

Corollary 3.7.3 *In the joined graph $\mathcal{H}$, formed by a joining set of Markov equivalent maximal ancestral graphs, if U is a minimal j-connecting path between x and y given Z, then there is no sequence of distinct vertices $\langle Q_1, Q_2, \ldots, Q_n \rangle, n > 1$ such that for each pair of vertices $\{Q_i, Q_{i+1}\}, 1 \leq i < n$, Q_i is a shielded collider on the discriminating path for Q_{i+1} (on U) and Q_n is a shielded collider on the discriminating path for Q_1 (on U).*

Proof:

Suppose for a contradiction that such a vertex set $\langle Q_1, Q_2, \ldots, Q_n \rangle$ exists. By Lemma 3.7.1 $n \neq 2$ so $n > 2$. Without loss of generality, suppose Q_1 is to the right of Q_n on U. Let j be the highest index such that Q_j is to right of Q_1, if such a vertex exists; otherwise let $j = 1$. Note that by Corollary 3.7.1 $j \neq n - 1$ (where $X = Q_{n-1}$, $Y = Q_n$, $Z = Q_1$ and $n - 1 \neq 1$ since $n > 2$).

I will now show that Q_{j+1} is to the left of Q_n. If $j = 1$, then by the definition of j, Q_2 is to the left of Q_1, and by Corollary 3.7.1 Q_2 is not in $U(Q_n, Q_1)$ (where $X = Q_1$, $Y = Q_2$ and $Z = Q_n$) so Q_2 is to the right of Q_n. If $1 < j < n-1$ then by the definition of j, Q_{j+1} does not lie to the right of Q_1. By Corollary 3.7.1 $Q_{j+1} \neq Q_n$, and by Corollary 3.7.2 Q_{j+1} does not lie on $U(Q_n, Q_1)$ (where $A_i = Q_n$, $A_{r+1} = Q_{j+1}$, $A_{i+1} = Q_1$ and $A_r = Q_j$). Hence Q_{j+1} is to the left of Q_n.

I will now show that the remaining vertices $Q_{j+m}, 2 \leq m \leq n-j-1$ lie to the left of Q_n which leads to a contradiction.

Base Case:$(m = j+2)$ By Corollary 3.7.1 Q_{j+2} does not lie on $U(Q_{j+1}, Q_j)$ so $Q_{j+2} \neq Q_n$ doesn't lie(where $X = Q_{j+1}$, $Y = Q_j$, and $Z = Q_{j+2}$).

Inductive Case:$(m = j+k)$ Assume for $1 < s < k$ Q_{j+s+1} lies to the left of Q_{j+s} and $U(Q_{j+s}, Q_{j+1})$ lies to the left of $U(Q_n, Q_1)$. By the definition of j, Q_{j+k+1} does not lie to the right of Q_1. By Corollary 3.7.1 Q_{j+k+1} does not lie on $U(Q_{j+k}, Q_{j+k-1})$ (where $X = Q_{j+k}$, $Y = Q_{j+k-1}$ and $Z = Q_{j+k+1}$). By repeated applications of Corollary 3.7.2 Q_{j+k+1} does not lie on $U(Q_{j+s}, Q_{j+s-1})$, $0 \leq s < k$; hence Q_{j+k+1} lies to the left of Q_{j+k}.

I have shown that each $Q_j, r \leq j \leq n$ lies on U in the following order: $\{Q_{n-1}, Q_{n-2}, \ldots, Q_{j+1}, Q_n, \ldots, Q_j\}$. But this is a contradiction by Corollary 3.7.2 (where $A_i = Q_{n-1}$, $A_{r+1} = Q_{j+1}$, $A_{i+1} = Q_n$ and $A_r = Q_j$). $\square$

Corollary 3.7.3 allows us to recursively define an order for a discriminating path for a variable on a minimal j-connecting path. Recall from Corollary 3.2.1 that if any two non-consecutive vertices on a minimal j-connecting path π given Z are adjacent, then the edge is directed and real, and the parent of this edge is a collider in Z. Thus, if $\langle \delta, \eta, \gamma \rangle$ forms a sub-path of π such that δ is adjacent to γ, then the edge is directed out of a collider (say γ) and there is some discriminating path for η (see Figure 3.38). So, I define the order of a discriminating path as follows:

Definition 3.7.1 *If $\langle x, B, y \rangle$ forms a sub-path of π, a minimal j-connecting path between α and β in a joined graph given Z, and B is unshielded then $\langle x, B, y \rangle$ is a 0^{th}-order shielded variable. If $U(x,y) = \langle x, q_1, \ldots, q_p, B, y \rangle$ forms a discriminating path for B (p can possibly equal 1) and no $q_i, 1 \leq i \leq p$ on $U(x,y)$ is shielded, then B is a 1^{st}-order shielded*

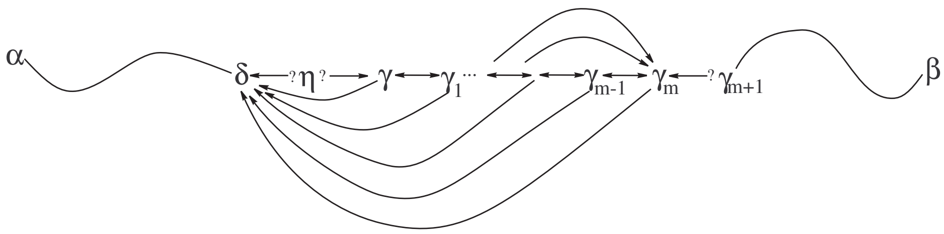


Figure 3.38: Diagram of a discriminating path with an order. See text for details.

variable. In general, if $U(x,y) = \langle x, q_1, \ldots, q_p, B, y\rangle$ is a discriminating path for B on π, then $order(B) = max\{order(q_i), 1 \leq i \leq p\} + 1$. i.e. if the maximum order of any q_i is $n-1$, then B is an n^{th}-order hidden variable. Furthermore, if $U(x,y)$ is a discriminating path for B and B is an n^{th}-order shielded variable, then $U(x,y)$ is said to be an n^{th}-order discriminating path for B.

Corollary 3.7.3 guarantees that that if π is a j-connecting path between α and β given Z in a joined graph, then there is at least one 0^{th}-order discriminating path for some vertex, and if any vertex on π is shielded then there is a 1^{st}-order shielded variable on π. Furthermore, the definition of the order of any shielded variable η on π is not defined in terms of the order of η itself. Lemma 3.7.2 shows that every shielded variable η on a minimal m-connecting path given Z, there is a unique sub-path of π that forms a discriminating path for η. Corollary 3.7.3 and Lemma 3.7.2 together guarantee that the definition of the order of any discriminating paths on π is well-defined.

Lemma 3.7.2 *Let π be a minimal j-connecting path between α and β given Z in the joined graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs. If $\alpha \ldots ? \rightarrow \gamma \leftarrow ?B? \rightarrow y$ is a sub-path of π and γ is a parent of y, then π contains a unique sub-path that forms a discriminating path for B.*

Proof:

Assume without loss of generality that $\pi = \langle \alpha, \ldots, x, \ldots, \gamma, B, y, \ldots, \beta\rangle$. I will show by induction that for each $n \geq 1$, if the following holds:

(i) π contains a vertex Q such that $\pi(Q,B)$ is of length n, and

(ii) every vertex q on $\pi(Q,B)$ except possibly B is a collider on $\pi(Q,B)$ and a parent of y in $\mathcal{H}$,

then π contains a vertex x such that $\pi(x,B)$ is of length $n+1$, and either $\pi(x,y)$ forms a discriminating path with order for B, or $x \to y$ is in $\mathcal{H}$ and x is a collider on π.

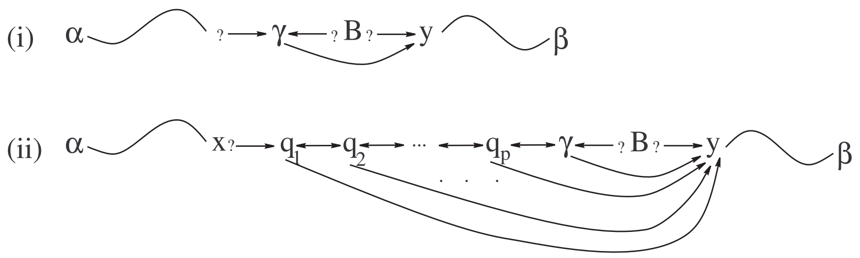


Figure 3.39: Unique sub-path of π forms a discriminating path for B in $\mathcal{H}$. See Lemma 3.7.2 for further explanation.

Base Case: By hypothesis, $\pi(\gamma,B)$ is a sub-path of π that is of length 1 such that (ii) holds (because there are no vertices between γ and B on π), and $\gamma \to y$ is in $\mathcal{H}$. See Figure 3.39(i).

Inductive Case: Assume that (i) and (ii) hold. If $Q = \alpha$, then the path $\langle \alpha = Q, y, \ldots, \beta \rangle$ is a j-connecting path given Z that violates the minimality of π. Thus, Q is not an endpoint, and there exists a vertex x to the left Q such that $\pi(x,B)$ is of length $n+1$. If $\pi(x,y)$ does not form a discriminating path for B, then x is adjacent to y. Furthermore, since $x? \to Q \to y$ is in $\mathcal{H}$, and by Corollary 3.2.1 the $x \to y$ edge is real. Again by Corollary 3.2.1, x is a collider on $\pi(x,B)$.

Arguing in this way, either there is a sub-path of π, U, that is shorter than the length of $\pi(\alpha,y)$ such that U forms a discriminating path for B, or every vertex between α and B is a collider on π, in which case $U = \pi(\alpha,y)$ forms a discriminating path for B and $x = \alpha$. See Figure 3.39(ii).

I will now show that U is unique. Because $\gamma \to y$ is in $\mathcal{H}$, no sub-path of $\pi(\gamma,\beta)$ can forms discriminating path for B and consequently x lies between α and y. Since every vertex between x and B on π is a parent of y, no sub-path of $\pi(x, y)$ forms a discriminating path for B. Finally, note that any sub-path $\pi(\delta, y)$ that contains $U = \pi(x, y)$ as a sub-path (i.e. δ is some vertex before x on $\pi(\alpha, x)$) does not form a discriminating path for B because x is not adjacent to y. Therefore U is unique. □

Lemma 3.7.3 *For every non-endpoint vertex on a minimal j-connecting path π between α and β given Z in the joined graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs there exists some discriminating path with an order.*

Proof:

By definition all unshielded vertices have an order. Consider any shielded vertex η, where $\langle \gamma, \eta, \delta \rangle$ is a sub-path of π with γ and δ adjacent. Suppose for a contradiction that η is a shielded vertex with no order. By Lemma 3.7.2 there is a discriminating path for η. If all vertices on the discriminating path have order then η has an order. Consequently, there exists at least one collider q_1 on the path without order. This vertex must be shielded (by definition of order) and thus by Lemma 3.7.2 there is a discriminating path for q_1. By Corollary 3.7.2 η is not on this discriminating path for q_1. Arguing in this way, we may construct an infinite sequence of shielded colliders $\{q_1, q_2, \ldots, q_n\}$ such that each q_i is a shielded vertex with no order and there is a discriminating path for q_i that does not contain $\{\eta, q_1, \ldots, q_{i-1}\}$. But this leads to a contradiction because there are only finitely many vertices on the path π. Thus, for sufficiently large n we exhaust the vertices on π such that no remaining vertices on π occur on the path for q_n.

Thus every non-endpoint vertex on π has an order and consequently, every non-endpoint vertex is discriminated by some path with order. □

3.8 Main Result 2: Global Markov Property

I now prove the main result of this chapter.

Theorem 3.8.1 *The vertices α and β are j-connected given Z in the joined graph $\mathcal{H}$ formed by joining a set of Markov equivalent maximal ancestral graphs if and only if α and β are* *j-connected given Z in every maximal ancestral graph joined to give rise to $\mathcal{H}$.*

In particular, I will show that every minimal j-connecting path given Z in $\mathcal{H}$ is m-connecting given Z in every maximal ancestral graph $\mathcal{G}$ joined to give rise to $\mathcal{H}$.

Proof:

($\Rightarrow$) Let π be a minimal j-connecting path between α and β given Z in $\mathcal{H}$. Suppose for a contradiction π is j-connecting given Z in $\mathcal{H}$ but the corresponding path is not m-connecting given Z in some $\mathcal{G}$ that gave rise to $\mathcal{H}$. By Lemma 3.4.1 all non-colliders on π in $\mathcal{H}$ are also non-colliders in $\mathcal{G}$. Therefore, all non-collider sub-paths of π are j-connecting in $\mathcal{G}$. By the definition of the join operation, all colliders on π in $\mathcal{H}$ are also present in $\mathcal{G}$. Therefore, if π is not j-connecting given Z in $\mathcal{G}$, then there is at least one collider on π that is not an ancestor of Z in $\mathcal{H}$. But by Lemma 3.5.6 all collider sub-paths on π in $\mathcal{H}$ are j-connecting given Z in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$ which is a contradiction. Hence if π is a minimal j-connecting path given Z in $\mathcal{H}$, then π is a j-connecting path given Z in all $\mathcal{G}$ joined to give rise to $\mathcal{H}$.

($\Leftarrow$) Let π be a minimal j-connecting path between α and β given Z in some maximal ancestral graph $\mathcal{G}$ that was used to give rise to the joined graph $\mathcal{H}$. Suppose for a contradiction that π is not j-connecting given Z in $\mathcal{H}$. By Lemma 3.7.3 every vertex along π in $\mathcal{G}$ is discriminated by some path with an order. Recall that m-connection is the same as j-connection for ancestral graphs because no arrowheads meet undirected edges in ancestral graphs. By Theorem A.0.1, all discriminating paths with order are common among all $\mathcal{G}^*$ joined to give rise to $\mathcal{H}$. Consequently, by the definition of join every vertex along π in $\mathcal{H}$ is a collider on π if and only if it is a collider along π in $\mathcal{G}$. By Lemma A.0.3, for every collider γ along π, the ancestral path from γ to a vertex in Z in $\mathcal{G}$ forms an ancestral path in every graph Markov equivalent to $\mathcal{G}$; and hence forms an ancestral path in $\mathcal{H}$. Thus π is j-connecting given Z in $\mathcal{H}$ which is a contradiction. □

By Theorem 3.8.1, if $\mathcal{H}$ be a graph formed by joining a set of Markov equivalent maximal ancestral graphs, then $\mathcal{H}$ is Markov equivalent to every maximal ancestral graph that gave rise to $\mathcal{H}$. It follows as a Corollary to Theorem 3.8.1 that these joined graphs also follow a pairwise Markov property.

Corollary 3.8.1 *If $\mathcal{H}$ is the graph formed by joining sets of Markov equivalent maximal ancestral graphs, then for every pair of non-adjacent vertices $\{\alpha, \beta\}$, there exists some subset of the remaining vertices that j-separates α and β.*

Proof:

For a contradiction, assume that α and β are non-adjacent vertices in $\mathcal{H}$ for which there is no set that j-separates them given Z. Then for any given Z there exists some path in $\mathcal{H}$ that j-connects α and β given Z. By Theorem 3.8.1, if α and β are j-connected given Z by a path in $\mathcal{H}$, then α and β are m-connected given Z by the corresponding path in any $\mathcal{G}$ used to form $\mathcal{H}$. Hence α and β are m-connected given any subset Z in $\mathcal{G}$. By the maximality of the ancestral graphs joined to give rise to $\mathcal{H}$, α and β are adjacent in $\mathcal{G}$, and hence in $\mathcal{H}$ which is a contradiction. □

Therefore I have shown that graphs formed by joining sets of Markov equivalent maximal ancestral graphs follow both global and pairwise Markov properties.

Theorem 3.8.2 *(Markov Equivalence for Joined Graphs) Joined graphs $\mathcal{H}_1$ and $\mathcal{H}_2$ are Markov equivalent if and only if $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same adjacencies and the same discriminating paths with order.*

Proof:

($\Rightarrow$) We first show that $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same adjacencies. Suppose for a contradiction that $\mathcal{H}_1$ and $\mathcal{H}_2$ are Markov equivalent to each other, but there is an edge between α and β in $\mathcal{H}_1$ that is not in $\mathcal{H}_2$. Then trivially, there is no subset of vertices that can j-separate α and β in $\mathcal{H}_1$. But by maximality, there is some set that j-separates α and β in $\mathcal{H}_2$, which is a contradiction. So $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same adjacencies.

Suppose for a contradiction that π forms a discriminating path with order for B in $\mathcal{H}_1$, but π does not consist of the same sequence of edge orientations $\mathcal{H}_2$. Since π has order, by definition it occurs as a sub-path of some minimal j-connecting path given Z in $\mathcal{H}_1$. By Corollary 3.7.3 every non-endpoint along π is discriminated by a path with order. By

Corollary A.0.2 π forms a discriminating path with order in $\mathcal{H}_2$ and B is a collider in $\mathcal{H}_2$ if and only if B is a collider in $\mathcal{H}_1$. So $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same discriminating paths with order.

($\Leftarrow$) Suppose for a contradiction that $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same adjacencies and the same discriminating paths with order, but $\mathcal{H}_1$ and $\mathcal{H}_2$ are not Markov equivalent. Let π be a minimal j-connecting path given Z between α and β in $\mathcal{H}_1$. We will show that π is j-connecting given Z in $\mathcal{H}_2$. By Corollary 3.7.3 every non-endpoint of π is discriminated by a path with order. By assumption $\mathcal{H}_1$ and $\mathcal{H}_2$ have the same discriminating paths with order thus π consists of the same sequence of edge orientations in $\mathcal{H}_2$ as in $\mathcal{H}_1$. Hence γ is a collider on π in $\mathcal{H}_2$ if and only if it is a collider on π in $\mathcal{H}_1$. Since π is j-connecting given Z in $\mathcal{H}_1$, γ is an ancestor of Z in $\mathcal{H}_1$. By Lemma 3.5.6 the ancestral path from γ to Z in $\mathcal{H}_1$ forms an ancestral path in $\mathcal{H}_2$. Hence π is j-connecting given Z in $\mathcal{H}_2$.

It has been shown that if π is a minimal j-connecting path given Z in $\mathcal{H}_1$, then π is also j-connecting given Z in $\mathcal{H}_2$. Therefore $\mathcal{H}_1$ and $\mathcal{H}_2$ are Markov equivalent to each other if and only if they have the same adjacencies and discriminating paths with order. $\square$

Chapter 4

CONCLUSIONS AND FUTURE WORK

Model search is an important part of statistical analyses, particularly in the absence of background knowledge. Joined graphs provide a way of graphically representing how much can be learned from data in the absence of background knowledge. However, even in the presence of background knowledge, processes can be represented by joined graphs that are consistent with the background knowledge. The following questions suggest directions for future research.

4.1 More on sup$[\mathcal{G}]$

Question 4.1.1 *What additional information can we gain by exploiting the fact that we have joined an entire equivalence class? For example, can directed cycles occur in sup*$[\mathcal{G}]$*?*

The results proven in Chapters 2 and 3 provide many insights into the structure of joined graphs. Throughout these chapters, at no time was it required that an entire equivalence class of maximal ancestral graphs be joined. Specifically, the main results, Theorems 2.6.1 and 3.8.1, and Corollary 3.8.1 hold for graphs formed by joining subsets of ancestral graphs. The structure of minimal j-connecting paths given Z is also preserved through the join operation.

Question 4.1.2 *What are the orientation rules for constructing sup*$[\mathcal{G}]$*?*

Meek (1995b), Andersson et al. (1997) and Chickering (1995) independently developed methods for constructing a graphical representation of Markov equivalence classes for DAGs by applying a number of orientation rules to any DAG. Such rules for joined graphs would allow one to construct sup$[\mathcal{G}]$ given any member of the equivalence class of joined graphs.

Establishing orientation rules for joined graphs may also provide insights into the following two questions:

Question 4.1.3 *When is* $\sup[\mathcal{G}]$ *an ancestral graph or a member of the equivalence class?*

Question 4.1.4 *When is* $\sup[\mathcal{G}]$ *an essential graph?*

The answer to these questions is work in progress.

4.1.1 Tails in $\sup[\mathcal{G}]$

My formulation of joined graphs draws no distinction between tails in an ancestral graph that are present across the Markov equivalence class of ancestral graphs, and tails that are present in only a subset of the equivalence class. If we have a tail at one end of an edge in $\mathcal{G}_1 \vee \mathcal{G}_2$ then it is possible that the tail is present in both $\mathcal{G}_1$ and $\mathcal{G}_2$ or in only one of them.

If we wish to know which ancestor relations are common to $\mathcal{G}_1$ and $\mathcal{G}_2$ then we need to distinguish between these two cases. The partial ancestral graph $\mathbf{\Phi}(\mathcal{G})$ does this by using a circle as an edge endpoint, when the tail is not always present in all graphs in $[\mathcal{G}]$. Hence we wish to answer the following two questions:

Question 4.1.5 *For which directed edges* $\alpha \to \beta$ *in* $\sup[\mathcal{G}]$ *does there exist a graph* $\mathcal{G}' \in [\mathcal{G}]$ *such that* $\alpha \leftrightarrow \beta$ *in* $\mathcal{G}'$*?*

Question 4.1.6 *For which undirected edges* $\alpha - \beta$ *in* $\sup[\mathcal{G}]$ *does there exist*

(i) a graph $\mathcal{G}' \in [\mathcal{G}]$ *such that* $\alpha - \beta$ *in* $\mathcal{G}'$*?*

(ii) a graph $\mathcal{G}'' \in [\mathcal{G}]$ *such that* $\alpha \leftarrow \beta$ *in* $\mathcal{G}''$*?*

(iii) a graph $\mathcal{G}''' \in [\mathcal{G}]$ *such that* $\alpha \leftrightarrow \beta$ *in* $\mathcal{G}'''$*?*

Lemmas 2.6.5 and 3.5.5, and Corollary 3.2.1 provide a partial answer to this question, but a full answer to these questions is as yet unknown.

4.1.2 Characterizing Joined Graphs

The edge types (undirected, singly directed and bi-directed) are not the only features defining ancestral graphs: the presence of an arrowhead at x on an $\{x, y\}$ edge in an ancestral graph implies that x is not an ancestor of y in the graph. It is clear from the examples given in section 2.3 that $\sup[\mathcal{G}]$ is not ancestral.

Question 4.1.7 *What are the edge configuration restrictions that characterize joined graphs that form* $\sup[\mathcal{G}]$ *for some Markov equivalence class of maximal ancestral graphs?*

The results in this thesis lend naturally to conjectures about the structure of $\sup[\mathcal{G}]$. A related question of interest is the following:

Question 4.1.8 *Does there always exist some set of* $\mathcal{G}$*s which when joined will form a given* $\mathcal{H}$*?*

Once the structure of joined graphs is characterized, such a representation can facilitate the model search process: instead of considering all possible models over the variables in the graph, one could first search across equivalence classes and then restrict a model search to models within that equivalence class.

BIBLIOGRAPHY

Andersson, S., D. Madigan, and M. Perlman (2001). Alternative markov property for chain graphs. *Scandinavian Journal of Statistics 28*, 33–85.

Andersson, S. A., D. Madigan, and M. D. Perlman (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics 25*, 505–541.

Chickering, D. (1995). A transformational characterization of equivalent bayesian networks. In P. Besnard and S. Hanks (Eds.), *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*, San Mateo, CA, pp. 87–98. Morgan Kaufmann.

Dawid, A. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society, B 41*, 1–31.

Frydenberg, M. (1990). The chain graph Markov property. *Scandin. J. Statist. 17*, 333–353.

Geiger, D., D. Heckerman, H. King, and C. Meek (1999). On the geometry of DAG models with hidden variables. In D. Heckerman and J. Whittaker (Eds.), *Preliminary papers of the Seventh International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 76–85.

Hernán, M., B. Brumback, and J. Robins (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology 11(5)*, 561–570.

Lauritzen, S. L. and N. Wermuth (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist. 17*, 31–57.

Meek, C. (1995a). Causal inference and causal explanation with background knowledge. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, San Francisco, pp. 403–410. Morgan Kaufmann.

Meek, C. (1995b). Strong completeness and faithfulness in Bayesian networks. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, San Francisco, pp. 411–418. Morgan Kaufmann.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann.

Pearl, J. (2000). *Causality: models, reasoning and inference.* Cambridge, UK: Cambridge University Press.

Richardson, T. and P. Spirtes (2000). Ancestral graph Markov models. To appear in Annals of Statistics.

Richardson, T. and P. Spirtes (To appear in 2002). Causal inference via ancestral graphs. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems.* Oxford University Press.

Settimi, R. and J. Q. Smith (1999). Geometry, moments and Bayesian networks with hidden variables. In D. Heckerman and J. Whittaker (Eds.), *Preliminary papers of the Seventh International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 293–298.

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

Spirtes, P., C. Meek, and T. Richardson (1995). Causal inference in the presence of latent variables and selection bias. In P. Besnard and S. Hanks (Eds.), *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, San Francisco, pp. 403–410. Morgan Kaufmann.

Spirtes, P., C. Meek, and T. Richardson (1999). Causal inference in the presence of latent variables and selection bias. In C. Glymour and G. F. Cooper (Eds.), *Computation, Causality and Discovery*, pp. 211–252. Cambridge, Mass.: MIT Press.

Spirtes, P. and T. S. Richardson (1997). A polynomial-time algorithm for determining DAG equivalence in the presence of latent variables and selection bias. In D. Madigan and P. Smyth (Eds.), *Preliminary papers of the Sixth International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pp. 489–501.

Verma, T. and J. Pearl (1991). Equivalence and synthesis of causal models. Technical Report R-150, Cognitive Systems Laboratory, UCLA.

Appendix A

MARKOV EQUIVALENCE OF MAXIMAL ANCESTRAL GRAPHS

Spirtes and Richardson (1997) state and prove conditions under which two maximal ancestral graphs are Markov equivalent to each other. Here we state and prove the same result under weaker conditions than required in Theorem 2.1.1. In particular, we weaken the condition of Spirtes and Richardson (1997) to require that the maximal ancestral graphs need only have the same adjacencies and discriminating paths with order to determine Markov equivalence. The definition of the order of a shielded vertex on a path for joined graphs (see Definition 3.7.1) is applicable to maximal ancestral graphs from the perspective that a maximal ancestral graph is a graph joined with itself, and recalling that j-connection is equivalent to m-connection for maximal ancestral graphs because arrowheads do not meet undirected edges in ancestral graphs. Furthermore, just as for joined graphs, the definition of discriminating paths with order is well-defined.

We now prove that every discriminating path with order on a minimal m-connecting path in a maximal ancestral graph $\mathcal{G}$ is also present in all graphs $\mathcal{G}^*$ Markov equivalent to $\mathcal{G}$, and that the vertex being discriminated is a collider in $\mathcal{G}$ if and only if it is a collider in $\mathcal{G}^*$.

Lemma A.0.1 *Let $U = \langle x, q_1, q_2, \ldots, q_p, B, y\rangle$ be a discriminating path with order for B in the maximal ancestral graph $\mathcal{G}$. If $\mathcal{G}^*$ is a maximal ancestral graph Markov equivalent to $\mathcal{G}$, and U forms a discriminating path for B in $\mathcal{G}^*$, and B is a collider on U in $\mathcal{G}$ if and only if B is a collider on U in $\mathcal{G}^*$.*

Proof:

Suppose for a contradiction that B is a collider in $\mathcal{G}$ but a non-collider in $\mathcal{G}^*$. Given a set Z, if Z does not contain all $q_i, 1 \leq i \leq p$, then the path $\langle x, q_1, \ldots, q_j, y\rangle$ is j-connecting where $q_j \notin Z$ and $q_i \in Z$ for all $i < j$. Since β is a collider on the path U in the graph

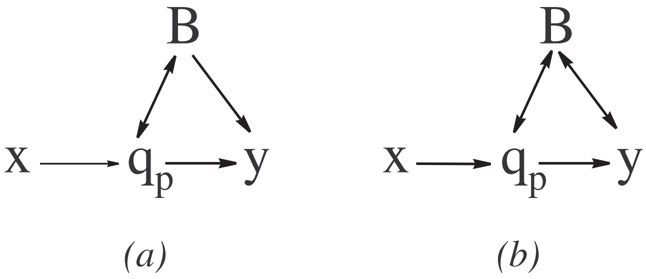


Figure A.1: Two ancestral graphs in which $U = \langle x, q_p, B, y\rangle$ is discriminating for B, but (a) and (b) are not Markov equivalent.

$\mathcal{G}$, if Z contains $\{q_1, \ldots, q_p\}$, then $\beta \notin Z$ if Z j-separates x and y. Consequently, in any graph Markov equivalent to $\mathcal{G}$ containing the discriminating path U, β is also a collider on U. Similarly, since β is a non-collider on the path U in $\mathcal{G}^*$, β is a member of any set that j-separates x and y, and β is a non-collider on U in any graph Markov equivalent to $\mathcal{G}^*$ containing U. But then we reach a contradiction because $\mathcal{G}$ and $\mathcal{G}^*$ do not encode the same set of conditional independence relations. See Figure A.1 $\square$

Corollary A.0.1 *(to Corollary 3.7.3) For every non-endpoint vertex on a minimal m-connecting path π between α and β given Z in a maximal ancestral graph $\mathcal{G}$ there exists some discriminating path with an order.*

Proof:

This proof immediately follows from Corollary 3.7.3 with $\mathcal{G}$ viewed as a graph joined with itself. $\square$

Corollary A.0.2 *Let $U = \langle x, q_1, q_2, \ldots, q_p, B, y\rangle$ be a discriminating path with order for B in the maximal ancestral graph $\mathcal{G}$. If U is a sub-path of a minimal m-connecting path π (π could possibly equal U), then U forms a discriminating path with order for B in any maximal ancestral graph that is Markov equivalent to $\mathcal{G}$ and the path U in $\mathcal{G}^*$ discriminates B to take the same form along π as it does on π in $\mathcal{G}$.*

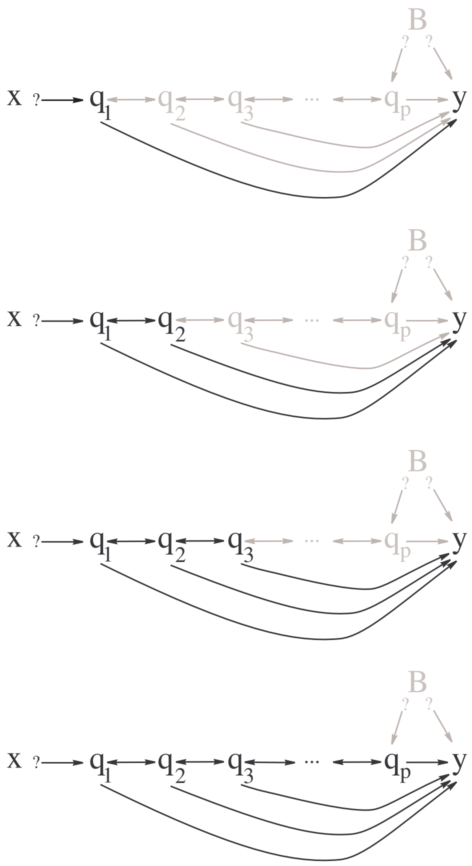


Figure A.2: Discriminating Path with order common to Markov equivalent maximal ancestral graphs. See Corollary A.0.2 for further details.

Proof:

Since U is discriminating for B, each vertex q_i $1 \leq i \leq p$ is a collider on U in $\mathcal{G}$. Since U is a sub-path of π, by Corollary A.0.1 every q_i is discriminated to be a collider along U in $\mathcal{G}$ and by Lemma A.0.1, every vertex q_i is also a collider on U in $\mathcal{G}^*$. See Figure A.2.

By the definition of a discriminating path, x is not adjacent to y, so $\{x, q_1, y\}$ forms an unshielded non-collider in $\mathcal{G}$, q_1 is discriminated to be a non-collider on this path, and by Lemma A.0.1, q_1 is a non-collider on this path in $\mathcal{G}^*$. Thus, $q_1 \rightarrow y$ occurs in $\mathcal{G}^*$. But then $\{x, q_1, q_2, y\}$ forms a discriminating path with order for q_2 in $\mathcal{G}$. By Lemma A.0.1 q_2 is a non-collider in $\mathcal{G}^*$ and consequently $q_2 \rightarrow y$ occurs in $\mathcal{G}^*$. By induction, each $q_i, 1 \leq i \leq p$ is discriminated by $\langle x, q_1, q_2, \ldots, q_i, y\rangle$ to be a non-collider on the path $\langle q_{i-1}, q_i, y\rangle$. By Lemma A.0.1 $\langle q_{i-1}, q_i, y\rangle$ forms a non-collider in $\mathcal{G}^*$ and hence $q_i \to y$ occurs in $\mathcal{G}^*$. Consequently, U forms a discriminating path with order for B in $\mathcal{G}^*$ and by Lemma A.0.1 B is a collider in $\mathcal{G}^*$ if and only if it is a collider in $\mathcal{G}$. □

Corollary A.0.3 *Let π be a minimal m-connecting path given Z in a maximal ancestral graph $\mathcal{G}$. If $\mathcal{G}^*$ is any maximal ancestral graph with the same adjacencies and discriminating paths with order, then γ is a collider along π in $\mathcal{G}$ if and only if γ is a collider along π in $\mathcal{G}^*$.*

Proof:

If γ is an unshielded collider along π, then γ is discriminated by a path with order 0. If γ is shielded, then by Corollary A.0.1 there exists some sub-path of π that forms a discriminating path with order for γ. By Corollary A.0.2 γ is a collider on π in $\mathcal{G}$ if and only if γ is a collider on π in $\mathcal{G}^*$. □

As a consequence of Corollary A.0.3, if $\mathcal{G}^*$ is Markov equivalent to $\mathcal{G}$, and π is a minimal m-connecting path given Z in $\mathcal{G}$, then each vertex along π is a collider in $\mathcal{G}^*$ if and only if it is a collider along π in $\mathcal{G}$. In other words, π consists of the same sequence of vertex types (collider or non-collider) in both $\mathcal{G}$ and $\mathcal{G}^*$. To show that π is m-connecting given Z in $\mathcal{G}^*$, it is necessary to show that every collider along the path is an ancestor of Z in $\mathcal{G}^*$. The following lemmas show just this, and are similar to lemmas proven by Richardson and Spirtes (2000).

Lemma A.0.2 *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two maximal ancestral graphs with the same adjacencies and discriminating paths with order. Let γ be a collider along π, a minimal m-connecting path given Z between vertices α and β in $\mathcal{G}_1$, and let η be the first vertex on the ancestral path from γ to a vertex in Z. Then the edge from γ to η is oriented $\gamma \to \eta$ in $\mathcal{G}_2$.*

Proof:

By Corollary A.0.1 every non-endpoint of π is discriminated by a path with order. By assumption $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same discriminating paths with order, so the sequence of edge orientations along π in $\mathcal{G}_2$ is the same as in $\mathcal{G}_1$. Therefore, γ is a collider in $\mathcal{G}_2$ if and only if it is a collider in $\mathcal{G}_1$.

Suppose for a contradiction that γ is not a parent of η in $\mathcal{G}_2$. There are two cases to consider:

(a) $\pi = \langle \alpha, \gamma, \beta \rangle$, $\alpha ?\!\rightarrow \gamma \leftarrow\!? \beta$ in $\mathcal{G}_1$ and $\mathcal{G}_2$.

Note that α is not adjacent to β or else the minimality of the path π would be violated. If α is not adjacent to η then $\langle \alpha, \gamma, \eta \rangle$ forms an unshielded non-collider and this path has order. Consequently, $\langle \alpha, \gamma, \eta \rangle$ forms a non-collider and $\gamma \rightarrow \eta$ in $\mathcal{G}_2$. Therefore α is adjacent to η, and similarly, β is adjacent to γ. Since $\mathcal{G}_1$ is ancestral, and $\alpha ?\!\rightarrow \gamma \rightarrow \eta$ is in $\mathcal{G}_1$, $\alpha ?\!\rightarrow \eta$ occurs in $\mathcal{G}_1$. Similarly, $\eta \leftarrow\!? \beta$ occurs in $\mathcal{G}_1$. But then $\langle \alpha, \eta, \beta \rangle$ forms an m-connecting path which violates the minimality of π. See Figure A.3. Since, if $\pi = \langle \alpha, \gamma, \beta \rangle$ then $\gamma \rightarrow \eta$ in $\mathcal{G}_2$, hence π is of length greater than 3.

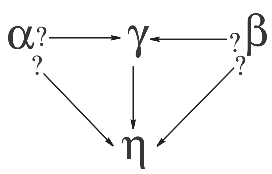


Figure A.3: If $n = 2$ and the path π is as outlined in Lemma A.0.2, then α and β are adjacent to η.

(b) $\pi = \langle \alpha, \ldots, x_{n+1}, x_n, \ldots, x_0, \gamma, y_0, \ldots, y_m, y_{m+1}, \ldots \beta \rangle$ where $n \geq 0,\ m \geq 0$.

If x_0 is not adjacent to η, then $\langle x_0, \gamma, \eta \rangle$ forms an unshielded non-collider in $\mathcal{G}_1$, and is a discriminating path with order 0. Then, since $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same discriminating paths with order, $\langle x_0, \gamma, \eta \rangle$ forms an unshielded non-collider and $\gamma \rightarrow \eta$ in $\mathcal{G}_2$, which is a contradiction. Therefore x_0 is adjacent to η. Similarly y_0 is adjacent to η. Because $\mathcal{G}_1$ is ancestral, and $x_0 ?\!\rightarrow \gamma \rightarrow \eta$ is in $\mathcal{G}_1$, $x_0 ?\!\rightarrow \eta$ occurs in $\mathcal{G}_1$. Similarly, $\eta \leftarrow\!? y_0$ occurs in $\mathcal{G}_1$.

There exists a vertex N between α and γ along π such that either:

(i) N is not adjacent to η,

(ii) N is a child of η, i.e. $N \leftarrow \eta$ is in $\mathcal{G}_1$, or

(iii) N is a collider along both π and π^* where $\pi^* = \langle \alpha, \ldots, N, \eta \rangle$, or N is a non-collider along both π and π^*.

Let x_{n+1} be the closest such vertex to γ (between α and β) where $\langle x_n, \ldots, x_0 \rangle$ are the vertices between x_{n+1} and γ along π. Note that by the definition of x_{n+1}, $\{x_n, \ldots, x_0\}$ are adjacent to η, parents of or spouses of η, and x_i is either a collider on π and a collider on π^* or vice versa for $0 \leq i \leq n$. Similarly, there exists a vertex M between γ and β along π such that either:

(i*) M is not adjacent to η,

(ii*) M is a child of η, i.e. $\eta \rightarrow M$ is in $\mathcal{G}_1$, or

(iii*) M is a collider along both π and π^* where $\pi^* = \langle \eta, M, \ldots, \beta \rangle$, or M is a non-collider along both π and π^*.

Let y_{m+1} be the closest such vertex to γ where $\langle y_0, \ldots, y_m \rangle$ are the vertices between γ and y_{m+1} along π.

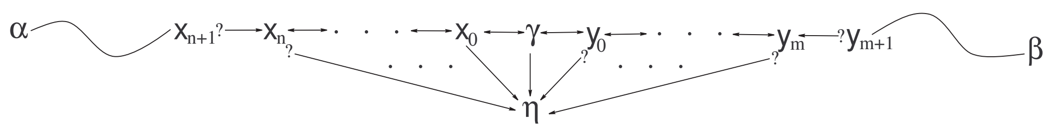


Figure A.4: If $n \geq 3$ and the path π is as outlined in Lemma A.0.2, then either $\langle \alpha, \ldots, x_{n+1}, \eta, y_{m+1}, \ldots, \beta \rangle$ is a more minimal j-connecting path given Z, or there is a discriminating path with order for γ that implies $\gamma \rightarrow \eta$ in all Markov equivalent maximal ancestral graphs. μ_0 and μ_2 are adjacent to α_0.

We will show by induction that for $0 \leq i \leq n$, x_i is a collider along π and a parent of η in $\mathcal{G}_1$, where $x_{-1} \equiv \gamma$. If $x_{n+1} \equiv x_0$, then it is trivially true because there are no edges between x_{n+1} and γ. Consider the case $n + 1 = 1$: we have already shown that x_0 is adjacent to η such that $x_0 ? \rightarrow \eta$. Therefore, by the definition of $x_{n+1} = x_0$, either $\eta \leftrightarrow x_0 \rightarrow \gamma \rightarrow \eta$ or $\eta \leftarrow x_0 \leftrightarrow \gamma \rightarrow \eta$ is in $\mathcal{G}_1$. Since $\mathcal{G}_1$ is ancestral, $\eta \leftarrow x_0 \leftrightarrow \gamma \rightarrow \eta$ occurs in $\mathcal{G}_1$. Assume for some $0 < k < n$ that x_k is such that $x_{k+1} ? \rightarrow x_k \leftrightarrow x_{k-1}$ and $x_k \rightarrow \eta$ occur in $\mathcal{G}_1$. Then by the definition of x_{n+1}, x_{k+1} is adjacent to η such that $x_{k+1} ? \rightarrow \eta$, and either $\eta \leftrightarrow x_{k+1} \rightarrow x_k \rightarrow \eta$ or $\eta \leftarrow x_{k+1} \leftrightarrow x_k \rightarrow \eta$ is in $\mathcal{G}_1$. Because $\mathcal{G}_1$ is ancestral, $\eta \leftarrow x_{k+1} \leftrightarrow x_k \rightarrow \eta$ occurs in $\mathcal{G}_1$, and by the definition of x_{n+1}, x_{k+1} is a collider on π but a non-collider on $\langle x_{k+2}, x_{k+1}, \eta \rangle$. Hence each of $\langle x_n, x_{n-1}, \ldots, x_0, \gamma\rangle$ is a collider on π and a parent of η in $\mathcal{G}_1$. Similarly, each of $\langle \gamma, y_0, \ldots, y_m\rangle$ is a collider on π and a parent of η in $\mathcal{G}_1$. See Figure A.4.

We now consider $n+1 > 1$. If x_{n+1} is not adjacent to η then $\langle x_{n+1}, x_n, \ldots, x_0, \gamma, \eta\rangle$ forms a discriminating path with order for γ which discriminates $\langle x_0, \gamma, \eta\rangle$ to be a non-collider. By assumption $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same discriminating paths with order so $\langle x_0, \gamma, \eta\rangle$ is a non-collider and $\gamma \rightarrow \eta$ occurs in $\mathcal{G}_2$, which is a contradiction. Therefore x_{n+1} is adjacent to η. Similarly, y_{m+1} is adjacent to η. Since (i) and (ii) are ruled out, $x_{n+1}? \rightarrow \eta \leftarrow ?y_{m+1}$ occurs in $\mathcal{G}_1$. Furthermore by the definition of x_{n+1}, either both $\langle x_{n+2}, x_{n+1}, x_n\rangle$ and $\langle x_{n+2}, x_{n+1}, \eta\rangle$ form colliders or both form non-colliders in $\mathcal{G}_1$. Similarly, either both $\langle y_m, y_{m+1}, y_{m+2}\rangle$ and $\langle \eta, y_{m+1}, y_{m+2}\rangle$ form colliders or both form non-colliders in $\mathcal{G}_1$. But then $\langle \alpha, \ldots, x_{n+1}, \eta, y_{m+1}, \ldots, \beta\rangle$ is a shorter m-connecting path given Z than π, which violates the minimality of π.

Since we reach a contradiction if γ is not a parent of η in $\mathcal{G}_2$, $\gamma \rightarrow \eta$ occurs in $\mathcal{G}_2$. $\square$

Lemma A.0.3 *Let $\mathcal{G}_1$ be a maximal ancestral graph and π be a minimal m-connecting path between α and β given Z in $\mathcal{G}_1$. If γ is a collider on π, then the shortest ancestral path from γ to Z is also ancestral in any maximal ancestral graph $\mathcal{G}_2$ that has the same adjacencies and the same discriminating paths with order as $\mathcal{G}_1$.*

Proof:

Let $\langle \gamma, \eta_0, \eta_1, \eta_2, \ldots, \eta_m\rangle$, where η_m is in Z, be the shortest ancestral path from γ to Z in $\mathcal{G}_1$.

By Lemma A.0.2 γ is a parent of η_0 in $\mathcal{G}_2$. It remains to be shown that all the subsequent edges on the ancestral path from γ to Z form an ancestral path in $\mathcal{G}_2$.

Suppose that $\eta_0 \rightarrow \eta_1$ does not occur in $\mathcal{G}_2$. Then $\gamma \rightarrow \eta_0 \rightarrow \eta_1$ is shielded since otherwise $\langle \gamma, \eta_0, \eta_1\rangle$ forms a discriminating path with order 0 and by assumption $\langle \gamma, \eta_0, \eta_1\rangle$ would form a non-collider in $\mathcal{G}_2$ such that $\eta_0 \rightarrow \eta_1$ occurs, which is a contradiction. By Lemma 2.5.3 $\gamma \rightarrow \eta_1$ is in $\mathcal{G}_1$, but then the minimality of the path is violated since we can take a directed path from γ to Z that bypasses η_0. So, $\eta_0 \rightarrow \eta_1$ occurs in $\mathcal{G}_2$. See Figure A.5.

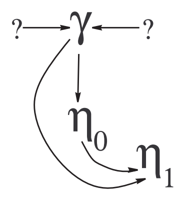


Figure A.5: If the path π is as outlined in the proof of Lemma A.0.3, then γ and η_1 are not adjacent such that $\gamma \to \eta_1$.

Assume that for $0 < k < m$, the $\eta_{k-1} \to \eta_k$ edge is real and consider the $\eta_k \to \eta_{k+1}$ edge. If the edge $\eta_k \to \eta_{k+1}$ is not present in $\mathcal{G}_2$, then η_{k-1} is adjacent to η_{k+1} (otherwise $\langle \eta_{k-1}, \eta_k, \eta_{k+1} \rangle$ forms an unshielded non-collider which by assumption would be present in $\mathcal{G}_2$, which is a contradiction). By Lemma 2.5.3 $\alpha_{k-1} \to \alpha_{k+1}$ is in $\mathcal{G}_1$, but then the minimality of the path is violated since we can take the path from γ to Z that bypasses η_k. Consequently, the $\eta_k \to \eta_{k+1}$ edge is present in $\mathcal{G}_2$. By induction, the edge orientations between $\eta_0, \eta_1, \eta_2, \ldots, \eta_m$ are present and form an ancestral path from η_0 to Z in $\mathcal{G}_2$.

Therefore, if γ is a collider on a minimal m-connecting path in the maximal ancestral graph $\mathcal{G}_1$, then the ancestral path from γ to Z forms an ancestral path in any maximal ancestral graph $\mathcal{G}_2$ that has the same adjacencies and discriminating paths with order as $\mathcal{G}_1$. $\square$

We now state and prove the main result of this section:

Theorem A.0.1 *(Markov Equivalence) Maximal ancestral graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies and the same discriminating paths with order.*

Proof:

($\Rightarrow$) We first show that $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies. Suppose for a contradiction that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent to each other, but there is an edge between α and β in $\mathcal{G}_1$ that is not in $\mathcal{G}_2$. Then trivially, there is no subset of vertices that can m-separate α and β in $\mathcal{G}_1$. But by maximality, there is some set that m-separates α and β in $\mathcal{G}_2$, which is a contradiction. So $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies.

Suppose for a contradiction that $\pi = \langle x, q_1, \ldots, q_p, B, y\rangle$ forms a discriminating path with order for B in $\mathcal{G}_1$, but π does not consist of the same sequence of edge orientations $\mathcal{G}_2$. Since π has order, by definition π is a minimal m-connecting path given $\langle q_1, \ldots, q_p\rangle$ if B is a non-collider, or $\langle q_1, \ldots, q_p, B\rangle$ if B is a collider in $\mathcal{G}_1$. By Corollary A.0.1 every non-endpoint along π is discriminated by a path with order. By Corollary A.0.2 π forms a discriminating path with order in $\mathcal{G}_2$ and B is a collider in $\mathcal{G}_2$ if and only if B is a collider in $\mathcal{G}_1$. So $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same discriminating paths with order.

($\Leftarrow$) Suppose for a contradiction that $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same adjacencies and the same discriminating paths with order, but $\mathcal{G}_1$ and $\mathcal{G}_2$ are not Markov equivalent. Let π be a minimal m-connecting path given Z between α and β in $\mathcal{G}_1$. We will show that π is m-connecting given Z in $\mathcal{G}_2$. By Corollary A.0.1 every non-endpoint of π is discriminated by a path with order. By assumption $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same discriminating paths with order thus π consists of the same sequence of edge orientations in $\mathcal{G}_2$ as in $\mathcal{G}_1$. Hence γ is a collider on π in $\mathcal{G}_2$ if and only if it is a collider on π in $\mathcal{G}_1$. Since π is m-connecting given Z in $\mathcal{G}_1$, γ is an ancestor of Z in $\mathcal{G}_1$. By Lemma A.0.3 the directed path from γ to Z in $\mathcal{G}_1$ forms a directed path in $\mathcal{G}_2$. Hence π is m-connecting given Z in $\mathcal{G}_2$.

It has been shown that if π is a minimal m-connecting path given Z in $\mathcal{G}_1$, then π is also m-connecting given Z in $\mathcal{G}_2$. Therefore $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent to each other if and only if they have the same adjacencies and discriminating paths with order. $\square$

VITA

Rebecca Ayesha Ali was born on December 15, 1974, in London, Ontario, Canada. In June, 1996 she received an Honours Bachelor of Science degree at the University of Western Ontario in Statistics and Actuarial Sciences. From 1996 to 1997 she worked as an Actuarial Consulting Assistant at Martineau Provencher (now AON) in Toronto, Ontario. She commenced studies at the University of Toronto in the fall of 1997, and graduated with a Masters of Science in Statistics in June 1998. In August 2002 she received her Doctor of Philosophy in Statistics from the University of Washington. She is currently working as an Assistant Professor in the Department of Statistics at the National University of Singapore.