02

Uncertainty in Artificial Intelligence

Proceedings of the Eighteenth Conference

Edited by

Adnan Darwiche

Nir Friedman

Uncertainty in Artificial Intelligence

Proceedings of the Eighteenth Conference (2002)

August 1–4, 2002

University of Alberta, Edmonton

Edited by

Adnan Darwiche, Univeristy of California, Los Angeles
Nir Friedman, Hebrew University, Jerusalem

Sponsored by

American Association for Artificial Intelligence
Boeing Research
Fair, Isaac and Company, Inc.
Hewlett-Packard Laboratories
Information Extraction and Transport, Inc. (IET)
Intel Labs
Microsoft Research
University of Alberta

Conference Chair

Daphne Koller
Stanford University

Morgan Kaufmann Publishers
San Francisco, California

Author Instructions, Electronic Design & Production, and Manufacturing by

Professional Book Center
2050 South Saint Paul Street
Denver, CO 80210
www.probook.net

Morgan Kaufmann Publishers
Editorial Office:
340 Pine Street, 6th Floor
San Francisco, California 94104

05 04 03 02 5 4 3 2 1

ISBN 1-55860-897-4

Uncertainty in Artificial Intelligence

Proceedings of the Eighteenth Conference (2002)

Preface

Since 1985, the Conference on Uncertainty in Artificial Intelligence (UAI) has been the primary international forum for presenting new results on the use of principled methods for reasoning under uncertainty within intelligent systems. The scope of UAI is wide, including but not limited to representation, automated reasoning, learning, decision making, and knowledge acquisition under uncertainty. The Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-2002) continues in this tradition, including contributions that report on advances in these core areas, as well as insights derived from the construction and use of applications involving uncertain reasoning.

This volume comprises the papers accepted for presentation at UAI-2002, held at the University of Alberta, Edmonton, Canada, from August 1 through August 4, 2002. Papers appearing in this volume were subjected to rigorous review—three Program Committee members (or in some cases, auxiliary reviewers) reviewed each paper under the supervision of an Area Chair, who made recommendations to the Program Chairs. This year 192 papers were submitted to UAI, with 66 accepted for presentation at the conference. All accepted papers appear in this volume.

Continuing with the tradition of UAI, two awards were presented at this year's conference: a Best Paper award for outstanding technical contribution and a Best Student Paper award. We are pleased to present the UAI-2002 Best Paper Award to Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky for their paper, "A New Class of Upper Bounds on the Log Partition Function." We are also pleased to present the UAI-2002 Best Student Paper award to Carlos Brito and Judea Pearl for their paper, "Generalized Instrumental Variables."

In addition to the presentation of technical papers, we are very pleased to have five distinguished invited speakers: Persi Diaconis (Stanford), Eric Grimson (MIT), Rob Schapire (AT&T), Sebastian Thrun (Carnegie Mellon University), and Peter P. Wakker (Maastricht University). Rob Schapire, Sebastian Thrun, and Peter P. Wakker have also contributed papers to this volume representing their invited talks. UAI-2002 also continues the tradition of offering a full-day course on Advanced Topics in Uncertainty, consisting of four tutorials by Craig Boutilier (University of Toronto), Michael Collins (AT&T), Dan Geiger (Technion), and Carla Gomes & Bart Selman (Cornell University).

The set of papers, invited talks, and full day course constitute the technical program of the UAI conference. We are proud of the quality of this year's program, and are looking forward to continued growth and contributions to future UAI conferences.

Adnan Darwiche and Nir Friedman
Program Co-Chairs

Daphne Koller
Conference Chair

Acknowledgments

The success of UAI-2002 and the quality of the proceedings are due in no small part to the tremendous efforts and talents of our Program Committee. Through their careful reviews of submitted papers and well-considered recommendations to the Program Chairs, they have ensured that the program and the proceedings reflect state-of-the-art research advances. Through their thoughtful feedback to the authors, they have helped improve the quality of the papers in this volume. This year the Program Committee was again structured to rely on 14 Area Chairs, who coordinated reviews and discussions from 78 Program Committee members and 57 additional reviewers. Area Chairs made recommendations to the Program Chairs on each paper based on the joint reviews and discussion. We are extremely grateful to the people below for their efforts in helping to make UAI-2002 a success.

Program Committee Area Chairs

Fahiem Bacchus, University of Toronto
Craig Boutilier, University of Toronto
Didier Dubois, Institut de Recherche en Informatique de Toulouse (IRIT)
Linda van der Gaag, Utrecht University
Dan Geiger, Technion
Lluís Godo, Institut d'Investigació en Intelligència Artificial (IIIA)
Eric Horvitz, Microsoft Research
Tommi Jaakkola, MIT
Michael Jordan, University of California, Berkeley
Uffe Kjaerulff, HUGIN
Chris Meek, Microsoft Research
Ron Parr, Duke University
David Poole, University of British Colombia
Peter Spirtes, Carnegie Mellon University

Program Committee

Hagai Attias
Salem Benferhat
Philippe Besnard
Kathryn Blackmond Laskey
Blai Bonet
Ronen Brafman
Jack Breese
Wray Buntine
Max Chickering
Michael Collins
Gregory Cooper
Robert Cowell
Fabio Cozman
James Cussens
Nando de Freitas
Bruce D'Ambrosio
Sanjoy Dasgupta
Tom Dean
Rina Dechter
Javier Diez
Marek Druzdzel
Gal Elidan
Brendan Frey
Hector Geffner
Lise Getoor
Zoubin Ghahramni
Bob Givan
Robert Goldman
Moises Goldszmidt
Amy Greenwald
Russell Greiner
Carlos Guestrin
Peter Haddawy
Petr Hajek
Joseph Halpern
Milos Hauskrecht
David Heckerman
Thomas Hofmann
Frank Jensen
Michael Kearns
Nevin L. Zhang
Jerome Lang
Helge Langseth
Steffen Lauritzen
Uri Lerner
Michael Littman
David Madigan
Anders Madsen
Marina Meila
Tom Minka
Paul-Andre Monney
Serafin Moral
Kevin Murphy
Andrew Ng
Thomas Nielsen
Kristian Olesen
Simon Parsons
Dana Pe'er
David Pennock
Avi Pfeffer
Henri Prade
Thomas Richardson
Irina Rish
Alessandro Saffiotti
Richard Scheines
Oliver Schulte
Dale Schuurmans
Ross Shachter
Prakash Shenoy
Yoav Shoham
Yoram Singer
Satinder Singh
Milan Studeny
Moshe Tennenholtz
Bo Thiesson
Jin Tian
Yair Weiss
Christopher Williams

We would also like to extend our appreciation to the following additional reviewers who assisted Program Committee members with the reviewing process: Jasmina Arifovic, Olav Bangsø, Chiranjib Bhattacharyya, Bozhena Bidyuk, David Blei, Hans Bodlaender, Andrés Cano, Hei Chan, Denver Dash, Sandip Debnath, James Delgrande, Dimiter Driankov, Alan Fern, Jeff Forbes, Lance Fortnow, Ya'akov Gal, José A. Gámez, Matthew Greig, Peter Hooper, Sandy Irani, Finn V. Jensen, Emir Kapanci, Lars Karlsson, Kalev Kask, Tomas Kocka, David Larkin, Kevin Leyton-Brown, Daniel Lizotte, Tsai-Ching Lu, Omid Madani, Robert Mateescu, Andrew McCallum, Yves Moinard, Iftach Nachman, Noam Nissan, Brenda Ng, Eugene Nudelman, Chris O'Brien, James Park, Mark Paskin, Jose M. Peña, Patrice Perny, Leonid Peshkin, Karen Pivazyan, Ryan Porter, Rob Powers, David Pynadath, Eric Raufaste, Dmitry Rusakov, Bikash Sabata, Eyal Shimony, Finnegan Southey, Masami Takikawa, Simon Tong, Ioannis Tsochantaridis, Jirka Vomlel, and Marta Vomlelova.

A note of thanks is due to the instructors in the full-day course on Advanced Topics in Uncertainty for putting together an outstanding program: Craig Boutilier, Michael Collins, Dan Geiger, and Carla Gomes & Bart Selman. We are also very grateful to our outstanding invited speakers, Eric Grimson, Rob Schapire, Sebastian Thrun, and Peter P. Wakker, as well as our banquet speaker, Persi Diaconis.

Special thanks are due to Russell Greiner of the University of Alberta for his invaluable assistance as the local arrangement chair. His help was crucial in making many aspects of the conference successful.

We owe a debt of gratitude to Microsoft Research, which hosted the reviewing process using their Conference Management Toolkit. We would also like to express appreciation to Jonathan Simon of Microsoft Research, who provided timely and responsive support to the Program Chairs in managing the conference review site. We also thank Jack Breese and the staff at Microsoft Research for their help in producing the registration materials.

Sincere thanks are also due to Information Extraction and Transport, Inc., which handled conference administration. We thank Christa Hopkins of IET for her cheerful and competent assistance with administrative matters. We also thank Jason Levitt for help in preparing and maintaining the web-based registration site.

Finally, we gratefully acknowledge the generous financial support to UAI-2002 by the following organizations: the American Association of Artificial Intelligence; Boeing Research; Fair, Isaac and Company, Inc.; Hewlett-Packard Laboratories; Information Extraction and Transport, Inc. (IET); Intel Labs; Microsoft Research; and the University of Alberta.

Contents

Markov Equivalence Classes for Maximal Ancestral Graphs

R. Ayesha Ali
Thomas S. Richardson
Department of Statistics
University of Washington
BOX 354322
Seattle, WA 98195-4322
ayesha@stat.washington.edu

Abstract

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. We introduce a simple representation of a Markov equivalence class of ancestral graphs, thereby facilitating the model search process for some given data. More specifically, we define a join operation on ancestral graphs which will associate a unique graph with an equivalence class. We also extend the separation criterion for ancestral graphs (which is an extension of d-separation) and provide a proof of the pairwise Markov property for joined ancestral graphs. Proving the pairwise Markov property is the first step towards developing a global Markov property for these graphs. The ultimate goal of this work is to obtain a full characterization of the structure of Markov equivalence classes for maximal ancestral graphs, thereby extending analogous results for DAGs given by Frydenberg (1990), Verma and Pearl (1991), Chickering (1995) and Andersson et al. (1997).

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

1 INTRODUCTION

A graphical Markov model is a set of distributions that can be described by a graph consisting of vertices and edges. The independence model associated with a graph is the set of conditional independence relations encoded by the graph. In this paper, we focus on the problem of learning causal structure. We suppose our observed data was generated by a process represented by a 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 some process represented by a given data-generating process $\mathcal{D}$ 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 we may not be able to determine the influence of any hidden variables, we 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: (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(i)*. 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

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

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(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: these models may not be identifiable; the likelihood of the parameters for a specific model may be multi-modal; inference may be highly sensitive to the assumptions made about the unobserved variables; and the associated distributions may be difficult to characterize, in particular they may not form a curved exponential family. See Settimi and Smith (1999) and Geiger et al. (1999).

If detailed background knowledge is known 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 $\mathcal{D}$ 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 $V \setminus (S \cup L)$ which is Markov equivalent to $\mathcal{D}$ on the $V \setminus (S \cup L)$ 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. We define a join operation on ancestral graphs which associates a unique graph with an equivalence class. We also extend the separation criterion (See Definition 2.2) for ancestral graphs (which is an extension of d-separation) and provide an outline of the proof of the pairwise Markov property for joined 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 (1995)). 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)).

Section 2 provides some basic definitions; Section 3 starts to characterize various aspects of joined graphs with respect to minimal inducing paths; Section 4 outlines the proof that the joined graph formed by joining Markov equivalent maximal ancestral graphs is itself maximal; and finally, Section 5 outlines areas for future research.

2 BASIC DEFINITIONS

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.

If α and β are vertices in a graph $\mathcal{G}$ such that $\alpha \leftrightarrow \beta$, then α is a *spouse* of β and vice versa.

If α and β are vertices in a graph $\mathcal{G}$ such that $\alpha \rightarrow \beta$, then α is a *parent* of β, and β is a *child* of α.

If there is a directed path from α to β (i.e. $\alpha \rightarrow\rightarrow \dots \rightarrow \beta$) or $\alpha = \beta$, 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.*

2.2 ANCESTRAL GRAPHS

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(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) provides 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 *A graph, which may contain undirected ($-$), directed edges ($\rightarrow$) 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.

A natural extension of Pearl's d-separation criterion may be applied to ancestral graphs. 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. $- v -$, $- v \rightarrow, \rightarrow v \rightarrow, \leftarrow v \rightarrow$.

Definition 2.2 *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;
(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 2.2 is an extension of the original definition of d-separation for DAGs in that the notions of 'collider' and 'non-collider' now include bi-directed and undirected edges. Since m-separation characterizes the independence relations in an underlying probability distribution compatible with a graph, tests of m-separation can be used to determine when graphs are Markov equivalent to each other.

Definition 2.3 *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(a), which motivates the following definition:

Definition 2.4 *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 2.5 *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.*

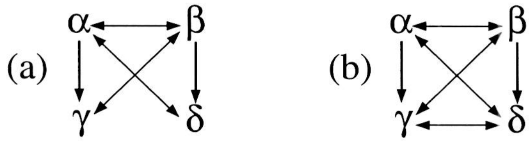

Figure 2: (a) The path $\{\gamma, \beta, \alpha, \delta\}$ is an example of an inducing path in an ancestral graph. (b) A maximal ancestral graph Markov equivalent to (a).

Figure 2(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(b).

Definition 2.6 *Suppose $\langle \alpha, \beta, \delta \rangle$ are vertices in a graph such that α and β are adjacent, and β and δ are adjacent. If α and δ are also adjacent, then $\langle \alpha, \beta, \delta \rangle$ is "shielded". Otherwise, if α and δ are not adjacent, then $\langle \alpha, \beta, \delta \rangle$ is "unshielded".*

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 $\mathcal{G}$; considering shielded colliders is not important in determining Markov equivalence for DAGs. *Discriminating paths* are useful for identifying which shielded colliders (and non-colliders) are required for ancestral graphs to be Markov equivalent:

Definition 2.7 *$U = \langle x, q_1, q_2, \ldots, q_p, \beta, y \rangle$ 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 contains $\{q_1, \ldots, q_p\}$ and β is a collider on the path U in the graph $\mathcal{G}$, then $\beta \notin Z$ if Z m-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, 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. 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. The paths $\langle x, q, \beta, y \rangle$ in $\mathcal{G}_1$ and $\mathcal{G}_2$ from Figure 4 are examples of discriminating paths for β. Note that if β is a non-collider on U, then $\beta \rightarrow y$ in $\mathcal{G}$.

Definition 2.8 *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 subpath of U from x to β forms a *collider path*. So referring to $\mathcal{G}_1$ and $\mathcal{G}_2$ in Figure 4, 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).

2.3 CHARACTERIZATION OF MARKOV EQUIVALENCE

Spirtes and Richardson (1997) proved the following result:

Theorem 2.1 *(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 in $\mathcal{G}_1$ if and only if it is a collider in $\mathcal{G}_2$.

2.4 JOINED GRAPHS

Here we define the join operation as a method of identifying the features 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 an OR operation on the "tails" of these graphs.

Definition 2.9 *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 $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$.

In general we 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 $\mathcal{G}$.

Figure 3 provides an example of a joined graph. Note that since there are arrowheads that meet the undi-

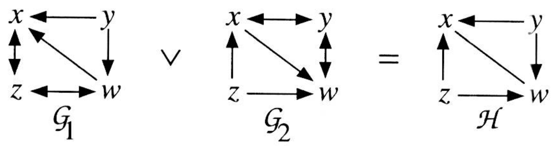

Figure 3: An example of joining two Markov equivalent ancestral graphs in which the joined graph is not ancestral.

rected edge $x - w$ in the joined graph, $\mathcal{H}$ is not ancestral as it violates condition *(c)* of Definition 2.1. Figure 4 shows another example of two Markov equivalent graphs being joined. Here, $\mathcal{H}$ is itself a member of the equivalence class of ancestral graphs.

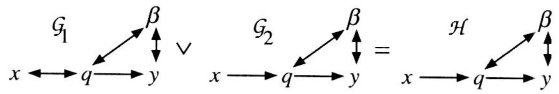

Figure 4: An example of joining two Markov equivalent ancestral graphs in which the joined graph is itself a member of the equivalence class.

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). Hence we restrict our attention to joining sets of Markov equivalent maximal ancestral graphs in the remainder of this paper. Ideally, any representation of an equivalence class of ancestral graphs would encode the same independence model encoded by all the ancestral graphs in the equivalence class.

We use the following notation 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 we only consider graphs with edges that are directed, bi-directed and undirected. By joining maximal ancestral graphs as outlined in Definition 2.9, the resulting joined graph $\mathcal{H}$ is not ancestral in general, see Figure 3. Given that undirected edges can meet arrowheads in joined graphs, what is the equivalent of a d-connecting path for joined graphs? Here we define a *j-connecting* path for joined graphs.

Definition 2.10 *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.

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

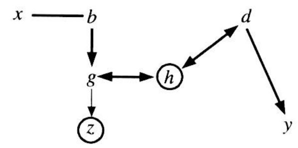

Figure 5: An example of a j-connecting path in a joined graph: x and y are j-connecting given $Z = \{z, h\}$.

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 3, we see that $\mathcal{H}$ 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 5 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 we extend the concept of maximality to joined graphs and in Section 4 we show that the graph $\mathcal{H}$ formed by joining Markov equivalent maximal ancestral graphs is itself maximal.

Definition 2.11 *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.

3 CHARACTERIZING THE JOINED GRAPH

To date, no full characterization of joined graphs is readily available. This section presents structural inferences that can be made about joined graphs. For instance, as with ancestral graphs, the configurations "$\to \gamma -$" and "$\leftrightarrow \gamma -$" do not occur in joined graphs. We 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.

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 3.1 *All bi-directed edges in a joined graph are real. Furthermore, if* $\alpha ? \to \beta - \gamma$ *in* $\mathcal{H}$*, then the* $\beta - \gamma$ *edge is not real.*

Proof: By the 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 if an undirected edge meets an arrowhead in a joined graph (using the example given in the proposition) then there is at least one ancestral graph that gave rise to $\mathcal{H}$ with an arrowhead at β on the $\beta - \gamma$ edge, i.e. the $\beta - \gamma$ edge is not real.

Lemma 3.2 *In a joined graph* $\mathcal{H}$*, formed by joining maximal ancestral graphs, if* $\gamma \to \beta ? - ?\delta ? \to \gamma$ *occurs and* $\gamma \to \beta$ *is real, then* $\beta \leftarrow ?\delta$ *also occurs in* $\mathcal{H}$*.*

Proof: There cannot be a tail at β on the $\{\beta, \delta\}$ edge in any ancestral graph that gave rise to $\mathcal{H}$ because in that case either $\delta \to \gamma \to \beta \to \delta$ or $\delta \leftrightarrow \gamma \to \beta \to \delta$ and the graph would not be ancestral. Thus, $\beta \leftarrow ?\delta$ in any graph $\mathcal{G}$ joined to form $\mathcal{H}$, and hence $\beta \leftarrow ?\delta$ in $\mathcal{H}$.

Lemma 3.3 *In a joined graph* $\mathcal{H}$*, formed by joining maximal 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. Then, β is not an ancestor of γ in any $\mathcal{G}$ that gave rise to $\mathcal{H}$. If the $\{\gamma, \delta\}$ edge is undirected, or there is an arrowhead at γ on this edge in $\mathcal{H}$, 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.

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.

3.1 INFERRING DISCRIMINATING PATHS

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

Lemma 3.4 *Let* $\mathcal{H}$ *be a graph formed by joining a number 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 2.7. Label the colliders on the path between x and β as $q_1, q_2, \ldots, q_p$, such that q_1 is adjacent to x, and q_p is adjacent to β. Note that $\langle x, q_1, q_2, \ldots, q_p, \beta\rangle$ 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 q_1 is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$. Thus, the $\{q_1, y\}$ edge in $\mathcal{H}$ is real. We 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 the $q_1 \to y$ edge is real, this discriminating path is present in all such $\mathcal{G}$, the $q_2 \to y$ edge in $\mathcal{H}$ is real. 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, $U = \langle x, q_1, q_2, \ldots, q_m, q_{m+1}, y\rangle$ discriminates $\langle q_m, q_{m+1}, y\rangle$ to be a non-collider in $\mathcal{H}$. Because each of $\{q_1, q_2, \ldots, q_m\}$ is a parent of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, U is a discriminating path present in all such $\mathcal{G}$ and hence the $\{q_{m+1}, y\}$ edge in $\mathcal{H}$ is real. Thus, by induction, $\langle q_1, q_2, \ldots, q_p\rangle$ are all parents of y in all $\mathcal{G}$ that gave rise to $\mathcal{H}$.

But then $U^* = \{x, q_1, q_2, \ldots, q_p, \beta, y\}$ forms a discriminating path for β in $\mathcal{H}$; U^* is present in all $\mathcal{G}$ that gave rise to $\mathcal{H}$, and thus $\langle q_p, \beta, y\rangle$ forms a collider in all $\mathcal{G}$ that gave rise to $\mathcal{H}$ if and only if $\langle q_p, \beta, y\rangle$ forms a collider in $\mathcal{H}$.

Corollary 3.1 *If a collider path* $q = \langle x, q_1, \ldots, q_p, \beta\rangle$ *is present in all Markov equivalent ancestral graphs*