

2-15-2023

Graphical Models in Reconstructability Analysis and Bayesian Networks

Marcus Andrew Harris
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Systems Science Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Harris, Marcus Andrew, "Graphical Models in Reconstructability Analysis and Bayesian Networks" (2023).
Dissertations and Theses. Paper 6336.
<https://doi.org/10.15760/etd.8190>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Graphical Models in Reconstructability Analysis and Bayesian Networks

by

Marcus Andrew Harris

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

Doctor of Philosophy
in
Systems Science

Dissertation Committee:
Martin Zwick, Chair
Jonathan Bird
Cliff Joslyn
Marek Perkowski

Portland State University
2023

Abstract

This research focuses on the investigation of two machine learning methodologies, Reconstructability Analysis (RA) and Bayesian Networks (BN). Both methods are probabilistic graphical modeling (PGM) methodologies. RA was developed in the systems community and has applications including time-series analysis, classification, decomposition, compression, pattern recognition, prediction, control, and decision analysis. BNs have origins in path models and have applications similar to those of RA. BNs are another graphical modeling approach for data modeling that is closely related to RA; where BN overlaps RA the two methods are equivalent, but RA and BN each has distinctive features absent in the other methodology.

The primary aim of this research is to make theoretical contributions through the unification of the RA and BN methods by developing and integrating the RA and BN neutral and directed system lattices and developing an algorithm to generate the joint RA-BN neutral system lattice of structures for any number of variables. This analysis is done exhaustively for four variables, which is sufficient to elucidate the formal relationship between these two PGM approaches.

The secondary aim of this research is to apply RA and BN to a real world problem in the electricity industry to identify predictive variables and obtain a new stand-alone model that improves prediction accuracy and reduces the INC and DEC Resource Sufficiency Requirements for Western Energy Imbalance Market participants.

The primary research aim was addressed by developing a lattice of structures for RA and BN that offers an expanded set of models to represent complex systems more accurately or more simply. The conceptualization of this lattice also offers a framework for additional innovations. Specific contributions include integrating RA and BN by developing and visualizing: (1) a BN neutral system lattice of general and specific graphs, (2) a joint RA-BN neutral system lattice of general and specific graphs, (3) an augmented RA directed system lattice of prediction graphs, and (4) a BN directed system lattice of prediction graphs. Additionally, it (5) extends RA notation to encompass BN graphs and (6) offers an algorithm to search the joint RA-BN neutral system lattice to find the best representation of the structure of the system variables. All lattices are for four variables, but the theory and methodology presented are general and apply to any number of variables. These methodological innovations are contributions to machine learning and artificial intelligence and more generally to complex systems analysis. These innovations also suggest extensions of RA and BN modeling that could enhance their power and flexibility.

The secondary aim was addressed by applying RA and BN, as well as Neural Networks and Support Vector Regression, to build predictive models of Net Load Imbalance for the Resource Sufficiency Flexible Ramping Requirement in the Western Energy Imbalance Market. This research identified predictive variables wind forecast, sunrise/sunset and the hour of day as primary predictors of net load imbalance, among

other variables, and show that the average size of the INC and DEC capacity requirements can be reduced by over 25% with the margin of error currently used in the industry while also significantly improving closeness and exceedance metrics. The reduction in INC and DEC capacity requirements would yield an approximate cost savings of \$4 million annually for one of nineteen Western Energy Imbalance market participants. Reconstructability Analysis performed best among the machine learning methods tested.

Acknowledgements

To my wife, Bryanna. Thank you for your unwavering support of my academic pursuits, and helping me stay the course even when work and life seemed to be taking over. You gave of your time and priorities to create space for me to pursue and complete it. For that, and many other reasons, I am forever grateful for your love and support.

To my children, Ella, Abigail, Jackson, and Lucy. Thank you for your ongoing interest in my academic work. Thank you for asking lots of questions and questioning whether what I was doing was 'really math'. Thank you also for your patience, it has felt like a long time to me, and I can imagine it seemed even longer to you. My hope for you in academics is simply to pursue things you are interested in and enjoy the process, even if it takes a long time. If you are interested and enjoy the process, your time will be well spent.

To Dr. Martin Zwick. Thank you for your many years of steadfast instruction and commitment to my success. Thank you for your patience to explain concepts until you were certain I had a sufficient understanding. Thank you for making the process stimulating and holding to the highest standards. You saw more in me than I saw in myself and drew the best out of me. That is a reflection of your world class academic instruction and of your incredible leadership. I would not be here without your guidance, instruction, and commitment over many years.

To my parents. Patrick and Janis Harris. Thank you for your unconditional love and support. Thank you for your genuine interest in everything that I do.

To Dr. Wayne Wakeland, Dr. Rajesh Venkatachalapathy, Dr. Joe Fusion, Dr. Alexandra Nielsen, Dr. Teresa Schmidt, Dr. Amanuel Melekin, and my other Systems Science peers. Thank you for creating a rich and creative learning environment in pursuit and advancement of the many facets of Systems Science.

To Elizabeth Kirby, and Francis Puyleart. Thank you for investing in me and passing on your extensive industry knowledge and expertise on balancing reserves and allowing me to try out some lesser heard of machine learning methods on your critical ancillary services issues.

To my committee Dr. Jonathan Bird, Dr. Cliff Joslyn, Dr. Marek Perkowski, and Dr. Martin Zwick. Thank you for your steadfast commitment to my academic pursuit and seeing me through to the finish.

Table of Contents

| | |
|--|-----------|
| ABSTRACT | i |
| ACKNOWLEDGEMENTS..... | iv |
| LIST OF TABLES..... | ix |
| LIST OF FIGURES | x |
| INTRODUCTION | 1 |
| CHAPTER 1 | 20 |
| PAPER 1 - JOINT LATTICE OF RECONSTRUCTABILITY ANALYSIS AND BAYESIAN NETWORK GENERAL | |
| GRAPHS..... | 20 |
| ABSTRACT | 21 |
| INTRODUCTION | 22 |
| RA LATTICE | 24 |
| BN LATTICE..... | 26 |
| Generating the lattice of BN general and specific graphs..... | 29 |
| Procedure to generate the BN general and specific graph lattice for any number of variables..... | 30 |
| Additional representations of BN general graph equivalence classes..... | 31 |
| INTEGRATING THE RHO, RA AND BN LATTICES | 31 |
| Lattice of four variable Rho graphs | 32 |
| Equivalent RA and BN general graphs..... | 33 |
| Non-equivalent RA and BN general graphs..... | 34 |
| JOINT LATTICE OF RA AND BN GENERAL GRAPHS..... | 35 |
| CONCLUSIONS..... | 36 |
| CHAPTER 2 | 38 |
| PAPER 2 - GRAPHICAL MODELS IN RECONSTRUCTABILITY ANALYSIS AND BAYESIAN NETWORKS | 38 |
| ABSTRACT | 39 |
| INTRODUCTION | 40 |
| RA LATTICE..... | 46 |
| RA Neutral Systems | 46 |
| Searching the RA Neutral System Lattice | 50 |
| RA DIRECTED SYSTEMS..... | 53 |
| Conventional Directed System Lattice | 53 |

| | |
|--|------------|
| Augmented Directed System Lattice | 57 |
| BN LATTICE | 61 |
| BN Introduction | 61 |
| BN Neutral Systems | 63 |
| BN Directed Systems | 81 |
| JOINT RA-BN NEUTRAL SYSTEM LATTICE | 85 |
| Joint RA-BN Neutral System Lattice Introduction | 85 |
| RA-BN Rho Neutral System Graphs | 86 |
| Rho and Equivalent RA and BN General Graphs | 88 |
| Rho and Non-Equivalent RA and BN General Graphs | 96 |
| Lattice of Rho, RA, BN Neutral System General Graphs..... | 97 |
| Joint RA-BN Lattice Algorithm | 98 |
| COMPARING RA AND BN DIRECTED SYSTEM GRAPHS | 102 |
| DISCUSSION..... | 106 |
| Neutral Systems..... | 106 |
| Directed Systems | 108 |
| CHAPTER 3 | 112 |
| PAPER 3 - MACHINE LEARNING PREDICTIONS OF ELECTRICITY CAPACITY..... | 112 |
| ABSTRACT | 113 |
| INTRODUCTION | 113 |
| MATERIALS AND METHODS | 119 |
| Data..... | 119 |
| METHODS..... | 124 |
| Industry Model | 125 |
| Reconstructability Analysis..... | 127 |
| Bayesian Networks | 132 |
| Support Vector Regression | 135 |
| Neural Networks..... | 136 |
| RESULTS OF DV PREDICTION: COMPARING ML METHODS TO INDUSTRY MODEL..... | 138 |
| Best point estimate predictions and comparison of methods..... | 138 |
| Best RA Model point estimate results | 140 |
| Best BN model point estimate results..... | 144 |
| BN Comparison to RA Best Model..... | 147 |
| SVM and NN point estimate results | 150 |
| RESULTS OF INC/DEC PREDICTION: COMPARING RA MODEL TO INDUSTRY MODEL..... | 151 |
| RA INC and DEC prediction procedure | 151 |
| Metrics for comparing INC and DEC predictions efficacy | 155 |
| Industry Model and RA Model INC and DEC predictions results | 161 |

| | |
|---|------------|
| RA Backup Model INC and DEC prediction results | 163 |
| DISCUSSION..... | 165 |
| CONCLUSIONS..... | 168 |
| AFTERWORD | 171 |
| SUMMARY OF ORIGINAL CONTRIBUTIONS..... | 171 |
| SUMMARY OF POSSIBLE RESEARCH EXTENSIONS | 176 |
| REFERENCES | 183 |
| APPENDICES..... | 189 |
| APPENDIX A.1 RA LOOP DETECTION PROCEDURE..... | 189 |
| APPENDIX A.2 D-SEPARATION PROCEDURE..... | 191 |
| Example 1 | 191 |
| Example 2 | 194 |
| APPENDIX B INPUT PARAMETERS FOR RA, BN, SVR, AND MLP ANALYSIS IN CHAPTER 3..... | 195 |

List of Tables

| | |
|--|-----|
| <i>Table 1 Example data for ABZ</i> | 8 |
| <i>Table 2 Calculated Probability Distributions A:B:Z</i> | 9 |
| <i>Table 3 Calculated Probability Distribution AZ:BZ</i> | 10 |
| <i>Table 4 Calculated Probability Distribution $ABZ_{A:B}$</i> | 11 |
| <i>Table 5 Calculated Probability Distribution AB from $ABZ_{A:B}$</i> | 11 |
| <i>Table 6 Calculated Probability Distribution AB:AZ:BZ</i> | 12 |
| <i>Table 7 Iterative Proportional Fitting Procedure Example for AB:AZ:BZ</i> | 13 |
| <i>Table 8 Model Transmission</i> | 15 |
| <i>Table 9 Rho, RA and BN equivalent graphs</i> | 34 |
| <i>Table 10 RA and BN terminology</i> | 66 |
| <i>Table 11 Four-variable independence statements</i> | 73 |
| <i>Table 12 Probability distribution and independencies of BN specific graph examples</i> | 80 |
| <i>Table 13 BN directed system graphs</i> | 82 |
| <i>Table 14 Equivalent Rho, RA and BN neutral system general graphs</i> | 95 |
| <i>Table 15 BN directed system graphs and RA equivalent example</i> | 104 |
| <i>Table 16 Variable Names and Definitions</i> | 120 |
| <i>Table 17 Cross Fold Validation</i> | 130 |
| <i>Table 18 $k = 12$ features used in SVR and NN models</i> | 137 |
| <i>Table 19 Results of Industry Model and Machine Learning Methods</i> | 139 |
| <i>Table 20 Best RA Model tested on all folds</i> | 141 |
| <i>Table 21 RA best model Independent Variables</i> | 142 |
| <i>Table 22 Rank order of variables for Fold 5 Best Overall Model</i> | 144 |
| <i>Table 23 Test of different prior link probabilities</i> | 147 |
| <i>Table 24 RA and BN Toy Example Results</i> | 149 |
| <i>Table 25 IVs found in Best RA, NN/SVR, RA Backup Models</i> | 151 |
| <i>Table 26 Observed Net Load Imbalance and INC and DEC predictions for each interval from Figure 46</i> | 159 |
| <i>Table 27 Calculated metrics from Figure 46 example</i> | 161 |
| <i>Table 28 Final Results</i> | 163 |
| <i>Table 29 RA Backup Model Point Estimate Summary Statistics</i> | 164 |
| <i>Table 30 RA Backup Model INC and DEC Predictions Compared to RA Model</i> | 165 |

List of Figures

Figure 1 RA and BN graph examples 7

Figure 2 RA General Graph Lattice..... 25

Figure 3 BN General Graph Lattice 28

Figure 4 Rho Lattice 32

Figure 5 RA G7 and BN2 general and specific graph example*..... 33

Figure 6 Joint Lattice of RA and BN General Graphs..... 37

Figure 7 Lattice of four-variable RA neutral system general graphs. 48

Figure 8 RA specific graph G15, AD:BD:CD 49

Figure 9 RA specific graph G13 50

Figure 10 Lattice of RA neutral system general and specific graphs 52

Figure 11 Conventional RA directed system lattice 56

Figure 12 Conventional RA directed system lattice and additional predictive specific graphs 57

Figure 13 Augmented RA directed system lattice..... 60

Figure 14 Lattice of BN neutral system general graphs..... 64

Figure 15 Examples of Markov equivalence tests..... 68

Figure 16 BN2, BN2b, BN2c*..... 69

Figure 17 PDAGs for graphs in Figure 14 insert 70

Figure 18 RA and BN notation example, without subscripts 71

Figure 19 BN notation examples with subscripts. (a) $BCD_{B:C}A$; (b) $ABCD_{AC:BC}$ 72

Figure 20 Lattice of general and specific BN neutral system graphs..... 76

Figure 21 Probability distribution for BN2 example*..... 80

Figure 22 BN directed system lattice 84

Figure 23 G15 and BN11 specific graph example* 86

Figure 24 Lattice of four-variable Rho graphs 87

Figure 25 Rho1, G1 and BN1 specific graph..... 88

Figure 26 Rho2, G7, and BN2 example* 89

Figure 27 Rho3, G10, and BN5 specific graph example* 90

Figure 28 Rho5, G13, and BN10 specific graph example* 91

Figure 29 Rho 6, G15 and BN11 specific graph example* 92

Figure 30 Rho7, G16 and BN14 specific graph example* 92

Figure 31 Rho 8, G17 and BN16 specific graph example* 93

Figure 32 Rho 9, G18 and BN18 specific graph example* 94

Figure 33 Rho, 10, G19 and BN19 specific graph example 94

Figure 34 Rho 11, G20 and BN20 specific graph..... 95

Figure 35 Lattice of 4-variable general Rho, RA and BN neutral system graphs..... 98

Figure 36 Example, Rho 2..... 99

Figure 37 Rho 2 example, with associated BNs general graphs 100

| | |
|--|------------|
| <i>Figure 38 RA and BN directed system lattice comparison</i> | <i>103</i> |
| <i>Figure 39 BN Directed System Prediction Example</i> | <i>106</i> |
| <i>Figure 40 Joint RA-BN lattice of 4 variable general and specific graphs</i> | <i>111</i> |
| <i>Figure 41 Industry Model calculation of INC and DEC requirement example.....</i> | <i>126</i> |
| <i>Figure 42 Best Overall BN</i> | <i>145</i> |
| <i>Figure 43 Best RA Model equivalent BN</i> | <i>148</i> |
| <i>Figure 44 RA and BN Toy Examples</i> | <i>150</i> |
| <i>Figure 45 RA calculation of INC and DEC for a given IV state.....</i> | <i>155</i> |
| <i>Figure 46 FRST Industry Model and RA Model Predictions Example</i> | <i>158</i> |
| <i>Figure 47 Test Results, Total Error and MW magnitude.....</i> | <i>163</i> |
| <i>Figure 48 Relationship of Wind Power Output to Wind Speed</i> | <i>167</i> |
| <i>193</i> | |

Introduction

Reconstructability Analysis (RA) and Bayesian Networks (BN) are machine learning (ML) methodologies that have many similarities. Both are probabilistic graphical modeling (PGM) methodologies that are qualitatively very different from neural networks, arguably the dominant machine learning approach in use today. In PGM methodologies, a model is a hypergraph that connects variables and a probability distribution defined on joint states of the variables. Although RA and BN are similar, no systematic examination has previously been made of the relationship between the two formalisms, and no comparison has previously been made of their relative efficacies as machine learning methods. This dissertation addresses both issues. It explores the formal relationship between the methods, and by showing where RA and BN overlap and give equivalent calculated distributions and also how RA and BN differ, it points to the possibility of a PGM approach that combines the unique features of both. This dissertation also reports a comparison of the relative performance of these two methods on Bonneville Power Administration data where predictive efficacy has potential practical value, while also comparing the performance of RA and BN to two other widely used ML methodologies.

RA was developed in the Systems community by Ashby (1964), Klir (1976, 1985, 1986), Conant (1981, 1988), Krippendorff (1981, 1986), Broekstra (1979), Cavallo (1979), Zwick (2001, 2004), and others. BNs have origins in path models originally described by Wright (1921, 1934), but it was not until the 1980s that BNs were more formally

established (Pearl, 1985, 1987, 1988; Neapolitan, 1989). This dissertation research centered on the theory and application of RA and BN and was developed through three published papers:

1. Joint Lattice of Reconstructability Analysis and Bayesian Network General Graphs¹
2. Graphical models in Reconstructability Analysis and Bayesian Networks²
3. Machine Learning Predictions of Electricity Capacity³

Papers 1 and 2 provide the primary theoretical contributions, unifying the RA and BN methodologies. Paper 2 builds on the findings in paper 1 and is really two papers fused into one: the first part develops the joint RA-BN lattice of *neutral* system general and specific graphs; the second part develops the RA *augmented directed* system lattice of general and specific graphs, among other new contributions. Paper 3 applies RA and BN to a real world industry problem to better predict electricity capacity requirements using these methods. Paper 3 also uses real and toy examples to further support and

¹ Harris, M.; Zwick, M. Joint Lattice of Reconstructability Analysis and Bayesian Network General Graphs. In Proceedings of the Tenth International Conference on Complex Systems; Springer: Cham, Switzerland, 2021.

² Harris, M.; Zwick, M. Graphical Models in Reconstructability Analysis and Bayesian Networks. *Entropy* 2021, 23, 986. <https://doi.org/10.3390/e23080986>

³ Harris M, Kirby E, Agrawal A, Pokharel R, Puyleart F, Zwick M. Machine Learning Predictions of Electricity Capacity. *Energies*. 2023; 16(1):187. <https://doi.org/10.3390/en16010187>

elaborate the theoretical findings in Papers 1 and 2. Paper 3 has two parts: the first part is an empirical comparison of RA and BN and two other more widely-used ML methods (neural networks and support vector machines); the second part applies the best performing ML method, which was RA, to industry data and shows that using RA predictive modeling a better solution can be obtained to an important industry problem than the industry approach currently used. The link between the first two papers and the third paper is theoretical. The first two papers point out that there are some models that are equivalent between RA and BN and some that are unique to each methodology. In the third paper it is found that the best RA model is in the group of RA models that are equivalent to some BNs which supports the finding from the first two papers.

RA is designed especially for nominal variables, but continuous variables can be accommodated if their values are discretized. (RA could in principle accommodate continuous variables, but this extension of the methodology has yet to be formalized.) Graph theory specifies the structure of the model: if the relations between the variables are all dyadic (pairwise), the structure is a graph; if some relations have higher ordinality, the structure is a hypergraph. Graph structures are independent of the data except for their needing to be supplemented by specification of variable cardinalities. In RA, information theory uses the data to characterize the precise nature and the strength

of the relations. Data applied to a graph structure yields a probabilistic graphical model of the data.⁴

RA can be applied to “neutral” and “directed” systems, and for both allows models that contain loops or do not contain loops. “Loops” in this dissertation refer to undirected graphs; “cycles” refer to directed graphs. Neutral systems characterize the relation among all variables, and applications are common in network analysis and image processing. Directed systems characterize the relationship between IVs and the DV. (In principle, RA could accommodate multiple DVs, but the specific implementation of RA used in this research (the Occam software package) allows only a single DV.⁵ For the data analysis work in this dissertation, RA directed systems are used because the primary goal is to predict the DV from the IVs.

Graphs are general or specific. A general graph identifies relations among variables that are unlabeled, i.e., variables whose identity is not specified; a specific graph labels (identifies) the variables. For example, for a system consisting of variables A, B, and C, AB:BC is a specific graph where nodes A and B are linked and B and C are also linked. Specific graphs AB:BC, BA:AC and AC:CB are all instances of the same general graph that has a unique independence structure regardless of variable labels. In this notation, the order of variables in any relation is arbitrary, as is the order of the relations. For example, CB:BA is identical to AB:BC. Relations include all of their

⁴ Paragraph from Harris and Zwick, 2021.

⁵ Paragraph from Harris, et. al., 2022.

embedded relations. For example, ABC includes embedded relations AB, AC and BC and the univariate margins A, B, and C. For directed systems, where the relationship of a single target variable (dependent variable) is the primary focus, the dependent variable is typically notated as “Z” and IVs are A, B, C and so on. The lattice of graphs for a neutral or a directed system with or without loops depends upon the number of variables in the data. For a three-variable neutral system allowing loops there are five general graphs and nine specific graphs; for four variables there are 20 general graphs and 114 specific graphs.⁶ As the number of variables increases further, the numbers of general graphs and specific graphs go up hyper-exponentially.

The lattice of four variable RA neutral systems is described in Chapter 2 “RA Lattice” and the four variable RA directed system lattice is also described in Chapter 2 “RA Directed Systems.” Here we define through a series of examples some of the key tenets of RA and BN structures, including marginal and conditional independence, as well as show how calculated probability distributions are derived for RA and BN graphs with unique independence structures for graphs with and without loops. These descriptions are related to information theoretic (probabilistic) RA, not set theoretic (crisp probabilistic) RA. Here we use directed systems; for neutral systems replace variable “Z” with variable “C” – the results would be the same in these examples.

⁶ Paragraph from Harris and Zwick, 2021.

The following RA and BN graph examples show representative structures with marginal or conditional independence, or complete dependence in the case of ABZ, their calculated probability distributions, and their independence structures. These examples illustrate some of the foundational elements of RA and BN graphs that will be expanded upon in much greater detail in Chapters 1 and 2. First we define three variables, A, B, Z to be used throughout the following examples. Variables are dichotomous with two states, 0 and 1.

Figure 1 shows a visual representation of the five examples to follow. These examples show RA and BN graphs that are equivalent, for example ABZ, A:B:Z, and AZ:BZ, a graph that is unique to RA, AB:AZ:BZ, that contains a loop, and a graph that is unique to BN, ABZ_{A:B}, that contains a V-structure. The examples that follow walk through these RA and BN structures, and use toy data to compute their calculated probability distributions in order to highlight some of the basic differences and similarities between certain graphs and the procedure for performing calculations when data is applied.


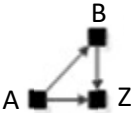
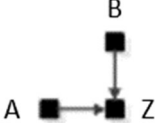
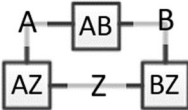
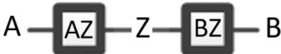
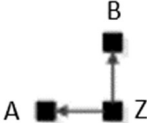
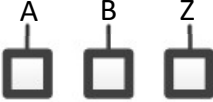
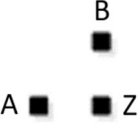
| | Model | RA Graph | BN Graph | Independence Statement | Probability Distribution |
|---|----------------|---|---|---|---|
| 1 | ABZ |  |  | n/a | $p(ABZ)$ |
| 2 | $ABZ_{A:B}$ | n/a (unique to BN) |  | $(A \perp B)$ | $p(Z AB)p(A)p(B)$ |
| 3 | $AB:AZ:B$ Z |  | n/a (unique to RA) | $(B \perp Z A),$ $(A \perp B Z),$ $(A \perp Z B)$ | n/a (no analytic solution, contains a loop) |
| 4 | $AZ:BZ$ |  |  | $(A \perp B Z)$ | $p(AZ)p(BZ)/p(Z)$ |
| 5 | $A:B:Z$ |  |  | $(A \perp B),$ $(A \perp Z),$ $(B \perp Z)$ | $p(A)p(B)p(Z)$ |

Figure 1 RA and BN graph examples

Table 1 shows example data for these three variables for RA graph ABZ (RA structures, which are hypergraphs, are referred to here using the more general term “graph.”). ABZ, shown below, is the raw data. To be more precise, usually the raw data is a contingency table of frequencies, and these frequencies when normalized by the sample size yield a probability distribution. The sample size is relevant for assessing the

statistical significance of deviations of a calculated model probability distribution from some reference distribution, but statistical considerations are not under discussion in this Introduction. All variables are dependent upon each other, therefore there are no independencies among variables. The probability (marginal value) of A0 is .6, which is the sum of the values in the p(ABZ) table in the first row of all four columns. The probability of B0 is .555, which is the sum of the values in the p(ABZ) table in the first and third columns. The probability of Z0 is .415, which is the sum of the values in the p(ABZ) table in the first two columns.

Table 1 Example data for ABZ

p(ABZ)

| | Z0 | | Z1 | |
|----|-------|-------|-------|-------|
| | B0 | B1 | B0 | B1 |
| A0 | 0.125 | 0.100 | 0.250 | 0.125 |
| A1 | 0.040 | 0.150 | 0.140 | 0.070 |

1.00

p(AB)

| | B0 | B1 | |
|----|-------|-------|-------|
| | A0 | 0.375 | 0.225 |
| A1 | 0.180 | 0.220 | 0.400 |
| | 0.555 | 0.445 | 1.000 |

p(Z)

| Z0 | Z1 |
|-------|-------|
| 0.415 | 0.585 |

1.000

If one were using the data as one's model, one would do prediction (classification), given values for the IVs, by generating and using the conditional probability distribution $p(Z|AB) = p(ABZ)/p(AB)$. One generally doesn't use the data as one's model, however, to avoid overfitting. Instead one selects a simpler model based on some selection criterion. One does prediction (classification) for this selected simpler

model in the same way: by using the conditional $q(Z|AB)$ generated from the joint calculated probability distribution, $q(ABZ)$, for the model.

The RA graph $A:B:Z$ represents marginal independence between all three variables, A, B, and Z. The colon, in RA notation, signifies marginal or conditional independence between variables – in this case marginal independence. That is, $(A \perp B)$, $(A \perp Z)$, $(B \perp Z)$ and these independence relations yield the calculated distribution $q(A:B:Z) = p(A)p(B)p(Z)$ shown in Table 2.

Table 2 Calculated Probability Distributions A:B:Z

$q(A:B:Z)$

| | Z0 | | Z1 | |
|----|-------|-------|-------|-------|
| | B0 | B1 | B0 | B1 |
| A0 | 0.138 | 0.111 | 0.195 | 0.156 |
| A1 | 0.092 | 0.074 | 0.130 | 0.104 |

1.00

For example, the first column and row (A0B0Z0) with value .138 is calculated by $p(A0)p(B0)p(Z0) = (.600)(.555)(.415) = .138$. A0B1Z1 is calculated by $p(A0)p(B1)p(Z1) = (.600)(.445)(.585) = .156$, and so on.

The RA graph $AZ:BZ$ represents *conditional* independence between A and B, given Z. That is, $(A \perp B | Z)$. This is a Naïve-Bayes-like graph structure and is common to both the RA and BN neutral system lattices, however it was not included in the RA directed system lattice until now. The RA augmented directed system lattice and the addition of this structure to this lattice are discussed in Chapter 2. The colon in the RA notation represents conditional independence between A and B given Z because Z is

included in the AZ term and the BZ term. This conditional independence relation yields the calculated distribution $q(AZ:BZ) = p(AZ)p(BZ)/p(Z)$ shown in Table 3. For example, the first column and row (A0B0Z0) with value .089 is calculated by $p(A0Z0)p(B0Z0)/p(Z0) = (.125+.100)(.125+.040)/(.415) = .089$.

Table 3 Calculated Probability Distribution AZ:BZ

q(AZ:BZ)

| | Z0 | | Z1 | |
|----|-------|-------|-------|-------|
| | B0 | B1 | B0 | B1 |
| A0 | 0.089 | 0.136 | 0.250 | 0.125 |
| A1 | 0.076 | 0.114 | 0.140 | 0.070 |

1.00

Related to the AZ:BZ structure is AB:BZ that is also found in the RA and BN lattices. AZ:BZ and AB:BZ both have the same general independence structure ($A \perp B \mid Z$), but given data produce different calculated probability distributions: for AZ:BZ, $q(AZ:BZ) = p(AZ)p(BZ)/p(Z)$ and for AB:BZ, $q(AB:BZ) = p(AB)p(BZ)/p(B)$.

The Bayesian Network graph $ABZ_{A:B}$ represents marginal independence between A and B within the triadic ABZ relation. That is, $(A \perp B)$. This independence relation yields the calculated distribution $q(ABZ_{A:B}) = p(Z \mid AB)p(A)p(B)$ where $p(Z \mid AB) = p(ABZ)/p(AB) = (p(ABZ)/p(AB)) * p(A)p(B)$. This BN contains what is called a “V-structure” where A is directed to Z and B is directed to Z. BNs with V-structures are unique to BN graphs and not found in RA and these V-structures result in a different factorization and thus calculated probability distribution than that of any RA graphs. While all non-V-structure BN graphs have an equivalent RA counterpart, all V-structure BNs are unique to BN. BN graphs and their relation to RA graphs are discussed in depth in Chapter 2,

section “BN Lattice.” The calculated distribution for $ABZ_{A:B}$ is shown in Table 4. For example, the first column and row (A0B0Z0) with value .111 is calculated by $(p(ABZ)/p(AB)) * p(A)p(B) = A0B0Z0/A0B0 * A0 * B0 = (.125/.375)(.6)(.555) = .111$.

Table 4 Calculated Probability Distribution $ABZ_{A:B}$

$q(ABZ_{A:B})$

| | Z0 | | Z1 | |
|----|-------|-------|-------|-------|
| | B0 | B1 | B0 | B1 |
| A0 | 0.111 | 0.119 | 0.222 | 0.148 |
| A1 | 0.049 | 0.121 | 0.173 | 0.057 |

1.00

That in this calculated distribution $q(ABZ_{A:B})$, A and B are marginally independent is shown by the AB distribution in Table 5. For example, A0B0 from Table 5 is generated from Table 4 by $A0B0Z0 (.111) + A0B0Z1 (.022) = .333$. The fact that $A0B0Z0 (qABZ_{A:B}) + A0B0Z1 (q ABZ_{A:B}) = A0 (pAB) * B0 (pAB)$ demonstrates that A and B are in fact marginally independent within the $q(ABZ_{A:B})$ calculated distribution $(.111 + .222 = .600 * .555)$.

Table 5 Calculated Probability Distribution AB from $ABZ_{A:B}$

$q(AB)$

| | B0 | B1 | |
|----|-------|-------|-------|
| A0 | 0.333 | 0.267 | 0.600 |
| A1 | 0.222 | 0.178 | 0.400 |
| | 0.555 | 0.445 | 1.000 |

RA graphs can have loops as illustrated by the RA graph $AB:AZ:BZ$ (independencies - $(B \perp Z | A)$, $(A \perp B | Z)$, $(A \perp Z | B)$). Whereas $A:B:Z$ and $AZ:BZ$ examples shown previously do not have loops, $AB:AZ:BZ$ has a loop. The loop within this structure does not depend on assigning any directed arrows to the relations AB, AZ, or BZ. (That

is, the order of the variables in these relations is arbitrary, as is also the order of the relations in this structure. Because the structure has a loop, the calculation of the AB:AZ:BZ probability distribution is not amenable to analytic solutions as in the prior examples. Rather, the calculated distribution for an RA graph with loops is done via iterative proportional fitting (IPF). An 'iteration' is defined here as the sequential imposition of all the relations in the graph, namely AB, AZ, and BZ. Table 6 shows the calculated distribution of AB:AZ:BZ using the $p(ABZ)$ data from Table 1.

Table 6 Calculated Probability Distribution AB:AZ:BZ

$q(AB:AZ:BZ)$

| | Z0 | | Z1 | |
|----|-------|-------|-------|-------|
| | B0 | B1 | B0 | B1 |
| A0 | 0.105 | 0.120 | 0.270 | 0.105 |
| A1 | 0.060 | 0.130 | 0.120 | 0.090 |

1.00

The calculated distribution for AB:AZ:BZ is generated by IPF. This procedure first starts with the uniform distribution for all states, i.e. .125. This can be seen in Table 7, Iteration 0 (the starting distribution). Then a series of iterations are performed imposing AB (Iteration 1.1), AZ (Iteration 1.2), and BZ (Iteration 1.3) distributions from $p(ABZ)$ in Table 1, then imposing AB again (iteration 2.1), AZ again (iteration 2.2), BZ again (iteration 2.3), and so on until the calculated distribution converges, where convergence is defined as the change in the probability distributions in successive iterations falling under some specified threshold.

For example, in order to impose the AB distribution onto the uniform distribution from Iteration 0. $(A0B0Z0(p\text{-Iteration } 0) * A0B0(pABZ)) / (A0B0Z0(p\text{-}$

Iteration 0) + A0B0Z1(p-Iteration 0)) = .125*(.125+.25)/(.125+.125) = .188 = A0B0Z0

Iteration 1.1. This procedure is continued for the remaining states of the ABZ

distribution in iteration 1.1. For Iteration 1.2, the p(AZ) distribution is imposed on the calculated distribution from Iteration 1.1.

Table 7 Iterative Proportional Fitting Procedure Example for AB:AZ:BZ

| p(ABZ) | | Z0 | | Z1 | | |
|--------------------------|----|-------|-------|-------|-------|------|
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.125 | 0.100 | 0.250 | 0.125 | 1.00 |
| | A1 | 0.040 | 0.150 | 0.140 | 0.070 | |
| Iteration 0 (uniform) | | Z0 | | Z1 | | |
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.125 | 0.125 | 0.125 | 0.125 | 1.00 |
| | A1 | 0.125 | 0.125 | 0.125 | 0.125 | |
| Iteration 1.1 p(AB) | | Z0 | | Z1 | | |
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.188 | 0.113 | 0.188 | 0.113 | 1.00 |
| | A1 | 0.090 | 0.110 | 0.090 | 0.110 | |
| Iteration 1.2 p(AZ) | | Z0 | | Z1 | | |
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.141 | 0.084 | 0.235 | 0.141 | 1.00 |
| | A1 | 0.086 | 0.105 | 0.094 | 0.115 | |
| Iteration 1.3 p(BZ) | | Z0 | | Z1 | | |
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.103 | 0.112 | 0.278 | 0.107 | 1.00 |
| | A1 | 0.062 | 0.138 | 0.112 | 0.088 | |
| Iteration 2.1 p(AB) | | Z0 | | Z1 | | |
| | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.101 | 0.115 | 0.274 | 0.110 | 1.00 |
| | A1 | 0.064 | 0.135 | 0.116 | 0.085 | |
| repeat until convergence | | Z0 | | Z1 | | |
| Iteration n | | B0 | B1 | B0 | B1 | |
| A | A0 | 0.105 | 0.120 | 0.270 | 0.105 | 1.00 |
| | A1 | 0.060 | 0.130 | 0.120 | 0.090 | |

Iterations continue until the calculated distribution converges or iteration might be halted upon reaching some preset maximum number of iterations. In this example, after 7 iterations, the calculated distribution was exactly the same as the prior iteration calculated distribution out to 10 significant digits.

All of the above examples in tables 2 through 7 represent RA or BN models with unique independence structures and thus unique calculated probability distributions of the data in Table 1. These models, along with the full set of four variable unique RA and BN models described in Chapter 2 convey different information about the data because each has a unique independence structure which represents the relationship between variables.

A variable relation is quantified by its uncertainty which in information theoretic RA is Shannon entropy. Shannon entropy for distribution $p(ABZ)$ is $H(ABZ) = -\sum p(ABZ) \log_2 p(ABZ)$ (Krippendorff, 1986). The difference between the uncertainty in one graph, given data, relative to another graph, represents information gain or loss; information loss relative to the data is model error known as Transmission. For the independence model, $H(A:B:Z) = H(A) + H(B) + H(Z)$, and Transmission $T(A:B:Z)$ is the uncertainty gain $H(A:B:Z) - H(ABZ)$, i.e., the error in model A:B:Z.

Table 8 shows Transmission in terms of information lost between the data $p(ABZ)$ and the example RA or BN graphs previously discussed. Here you can see that A:B:Z has the greatest information loss (error, uncertainty gain), which is as expected

because it represents complete variable independence. Conversely $ABZ_{A:B}$ represents the least information loss, which is also as expected as this graph structure contains only one independency ($A \perp B$).

Table 8 Model Transmission

| | $T(ABZ)$ | $T(ABZ_{A:B})$ | $T(AB:AZ:BZ)$ | $T(AB:BZ)$ | $T(AZ:BZ)$ | $T(A:B:Z)$ |
|---------------------|----------|----------------|---------------|------------|------------|------------|
| Transmission | - | 0.019 | 0.022 | 0.024 | 0.038 | 0.097 |

BNs have a similar interpretation for neutral system general and specific graphs; these are called in the literature “essential graphs” and “labeled essential graphs” respectively (Andersson, 1997), among other naming conventions. BN essential graphs, here called BN general graphs, can have multiple directed acyclic graph (DAG) representations while retaining the same unlabeled independence structure. For example, the BN $A \rightarrow B \rightarrow C$ has the same independence structure as $A \leftarrow B \rightarrow C$. In more complex structures with any number of variables, DAGs with different edge topologies can produce the same unlabeled independence structures, i.e. general graphs. This is discussed in Chapters 2 and 3, and has implications that may be initially unanticipated. For example, one might expect that a four variable maximally interconnected BN, which includes six directed edges, like BN 1 discussed in Chapter 1, would have six child graphs because a child graph results from the deletion of one edge from the parent graph. However this is not the case. BN1 has three child general graphs, not six, because deleting some of the edges from BN1 result in the same independence structure, and are thus redundant.

Prior to this research the BN literature did not have the concept of *directed* system general or specific graphs. In papers 1 and 2, BN essential graphs and labeled essential graphs are named using the RA notation for neutral system general and specific graphs, and BN directed system prediction graphs are developed for general and specific graph structures.

In Paper 1, RA neutral system general graphs are compared with BN neutral system general graphs, and a visualization of their lattice of graph structures for four variables is developed. This work shows that there are some RA general graphs with probability distributions unique to RA, there are some BN graphs with probability distributions unique to BN, and there are RA and BN graphs whose probability distributions are the same. This analysis also showed how these general graphs structures can be integrated into a unified lattice. This expands the set of general graphs with unique structures beyond what was previously available by either RA alone or BN alone, thus allowing for representations of complex systems which are (i) more accurate relative to data and/or (ii) simpler and thus more comprehensible and more generalizable than would be possible by modeling only with RA or only with BN.

Paper 1 is restricted to general graphs and to neutral systems. Paper 2 builds on the ideas from Paper 1 to expand the theoretical comparison between RA and BN by also considering specific graphs and directed systems. Paper 2's first new contribution is to develop a BN neutral system lattice of general and specific graphs and to compare that to the RA neutral system lattice of general and specific graphs. The BN four variable

lattice was developed by evaluating all 543 possible directed acyclic graphs for four variables, grouping graphs with the same unlabeled independence structures into 20 general graph equivalence classes, and identifying the 185 unique specific graphs associated with the 20 general graph equivalence classes. Paper 2 concretizes the general graph findings in Paper 1 by showing that when data is applied to RA or BN general graphs structures resulting in specific graph structures, some RA specific graphs have unique probability distributions, some BN specific graphs have unique probability distributions, and some RA and BN specific graphs are equivalent in their probability distributions.

Second, Paper 2 develops the four variable BN lattice of *directed* systems. In contrast to neutral systems, directed systems are focused, or 'directed', to a single variable of interest whereas neutral systems focus on the relationship among all variables. Directed systems are predominately used for predictions of a single dependent variable (DV), i.e., response variable, from independent variables (IVs), i.e., explanatory variables. In order to compare to RA directed systems, Paper 2 further expands the conventional RA directed system lattice to add "BN-like" -specifically Naïve-Bayes-like - graph structures that are found in the RA neutral systems lattice. Thus, Paper 2 expands the conventional RA directed system lattice creating a new augmented RA directed system lattice, and this lattice is compared and integrated with the BN directed system lattice also newly developed in this paper.

Third, in Paper 2, RA notation is extended to encompass BN graphs and an algorithm to generate the RA and BN joint lattice of general and specific graphs is developed.

Paper 3 applies real electrical system data to real and toy graph structures to give concrete examples of two theoretical findings in papers 1 and 2; specifically, that (i) there are multiple BN general graphs that are equivalent to a single RA general graph and that (ii) while the RA graph with the highest prediction efficacy on real data can be represented exactly by a BN graph, the BN search algorithm of a widely used software package is unable to find this graph within the BN lattice and thus would need to be improved to be able to do so.

In the application portion of this research in Paper 3, machine learning methods were used to make better predictions of the necessary electricity capacity requirement to participate in the California Independent System Operators Western Energy Imbalance Market (WEIM). Four methods are examined and compared: RA, BN, Support Vector Regression, and Neural Networks. These four methods are used to analyze data provided by the Bonneville Power Administration to build point estimate prediction models of observed net load imbalance, the dependent variable (DV.) These methods are compared to each other for prediction efficacy and compared to the current industry model. From this first step, the best performing method, which is RA, is used to develop a predictive model of the WEIMs Flexible Ramp Sufficiency Test (FRST) INC and DEC uncertainty requirement.

The findings in Paper 3 show that RA performs best among the four methods tested; the next best method is neural networks. The RA model more than doubles the R squared statistic relative to the industry model and reduces the size of the capacity requirement by 25.4%. This has real world significance in the amount of energy flexibility (capacity) that needs to be held by WEIM participants. Additionally, Paper 3 identifies highly predictive variables that increase prediction efficacy.

The following three chapters include the three published papers. The final chapter, the Afterword, includes a summary of the important findings and also of many possible extensions of this research.

Chapter 1

Paper 1 - Joint Lattice of Reconstructability Analysis and Bayesian Network General Graphs

Harris, M.; Zwick, M. Joint Lattice of Reconstructability Analysis and Bayesian Network General Graphs. In Proceedings of the Tenth International Conference on Complex Systems; Springer: Cham, Switzerland, 2021.

Co-authored by Marcus Harris and Martin Zwick.

Author contribution statement: Marcus Harris performed the conceptualization, formal analysis, writing original draft, and visualization. Dr. Martin Zwick performed conceptualization, writing review and editing, and supervision.

DOI: <https://doi.org/10.3390/e23080986>

Link to published paper: <https://archives.pdx.edu/ds/psu/33724>

Abstract

This paper integrates the structures considered in Reconstructability Analysis (RA) and those considered in Bayesian Networks (BN) into a joint lattice of probabilistic graphical models. This integration and associated lattice visualizations are done in this paper for four variables, but the approach can easily be expanded to more variables. The work builds on the RA work of Klir (1985), Krippendorff (1986), and Zwick (2001), and the BN work of Pearl (1985, 1987, 1988, 2000), Verma (1990), Heckerman (1994), Chickering (1995), Andersson (1997), and others. The RA four variable lattice and the BN four variable lattice partially overlap: there are ten unique RA general graphs, ten unique BN general graphs, and ten general graphs common to both RA and BN. For example, the specific graph having probability distribution $p(A)p(C)p(B|AC)$ is unique to BN, the RA specific graph $AB:AC:BC$, which contains a loop, is unique to RA, and the specific graph $ACD:BCD$ with probability distribution $p(A|CD)p(B|CD)p(D|C)p(C)$ is common to both RA and BN. The joint RA-BN lattice of general graphs presented in this paper expands the set of general graphs with unique independence structures beyond what was previously available by either RA alone or BN alone, thus allowing for representations of complex systems which are (i) more accurate relative to data and/or (ii) simpler and thus more comprehensible and more generalizable than would be possible by modeling only with RA or only with BN.

Introduction

Reconstructability Analysis (RA) is a data modeling approach developed in the systems community (Ashby, 1964; Klir, 1976, 1985, 1986; Conant, 1981, 1988; Krippendorff, 1981, 1986; Broekstra, 1979; Cavallo, 1979; Zwick, 2001, 2004; and others) that combines graph theory and information theory. Its applications are diverse, including time-series analysis, classification, decomposition, compression, pattern recognition, prediction, control, and decision analysis (Zwick, 2004). It is designed especially for nominal variables, but continuous variables can be accommodated if their values are discretized. Graph theory specifies the structure of the model (the set of relations between the variables), which is independent of actual data (except for specification of variable cardinalities); information theory uses the data to characterize the strength and the precise nature of the relations. Data applied to a graph structure yields a probabilistic graphical model of the data.

In RA, graphs are undirected, although directions are implicit if one variable is designated as the response variable (dependent variable), while all other variables are designated as explanatory variables (independent variables). In this paper, such IV-DV designations are not made, so we are concerned with only what are called “neutral systems.” Graphs are either general or specific. A general graph identifies relations among variables that are unlabeled; a specific graph labels the variables. For example, for a system consisting of variables A, B, and C, AB:BC is a specific graph where nodes A and B are linked and B and C are also linked. In this notation, the order of variables in

any relation is arbitrary, as is the order of the relations. Relations include all of their embedded relations. For example, the relation ABC includes embedded relations AB, AC and BC and the univariate margins A, B, and C. Specific graphs AB:BC, BA:AC and AC:CB are all instances of the same general graph that contains a unique independence structure regardless of variable labels.

For a three variable neutral system with loops there are 5 general graphs and 9 specific graphs; for four variables there are 20 general graphs and 114 specific graphs. The number of graphs increases hyper-exponentially with the number of variables.

A Bayesian Network (BN) is another graphical modeling approach for data modeling that is closely related to RA; indeed where it overlaps RA the two methods are equivalent, but RA and BN each has distinctive features absent in the other methodology. BNs have origins in the type of path model originally described by Wright (1921, 1934), but it was not until the 1980s that BNs were more formally established (Pearl, 1985, 1987, 1988; Neapolitan, 1989).

As does RA, BN combines graph theory and probability theory; graph theory provides the structure and probability theory characterizes the nature of relationships between variables. BNs are represented by a single type of graph structure; a directed acyclic graph, which is a subset of chain graphs, also known as block recursive models (Lauritzen, 1996). For a three variable BN lattice, there are 5 general graphs and 11 specific graphs; for four variables there are 20 general graphs and 185 specific graphs.

This paper integrates RA and BN and visualizes their joint lattice of general graphs for four variables.

RA Lattice

The lattice of four variable RA general graphs of Figure 2 (adapted from Klir, 1985; Krippendorff, 1986) represents all four variable RA graphs with unique independence structures. In these graphs, lines (which can branch) are variables; boxes are relations between variables. Where only two lines extend from a box, the relation is dyadic. If more than two lines extend from a box, the graph is a hypergraph. Bolded general graphs in Figure 2 are acyclic whereas non-bolded general graphs have cycles. The figure shows all 20 general graphs for four variables. G1 is the most complex general graph, in which all four variables are connected in a tetradic relation. Graphs below G1 are increasingly less complex decompositions of G1, ending with G20, the least complex graph, representing complete independence among all four variables.

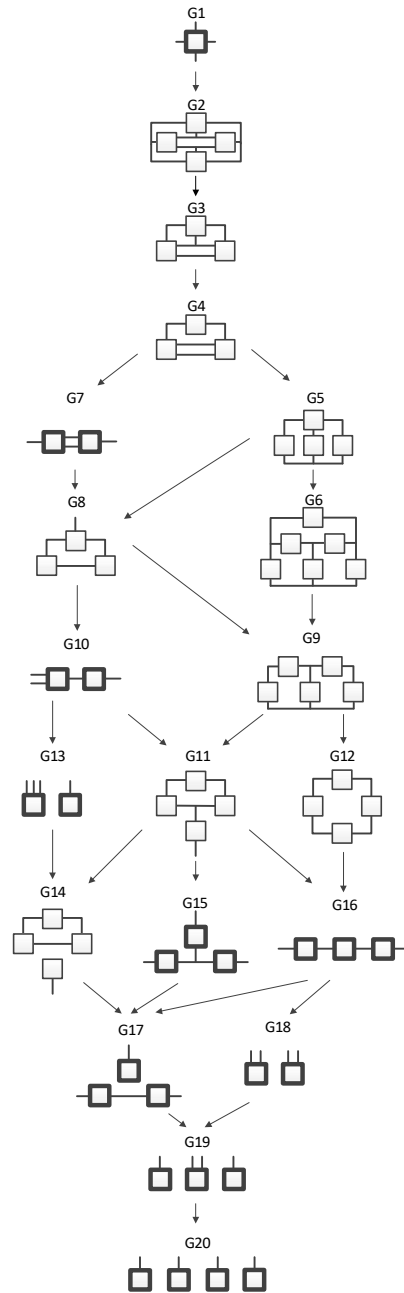


Figure 2 RA General Graph Lattice

A general graph represents a unique independence structure which disregards all possible ways that variables could be labeled. For example, the left-most and right-most variables in G7 in Figure 2 are independent of one another given the two central

variables that connect both relations, where these four variables have not been assigned any specific identities.

When, in RA or BN, the variables of a general graph are labeled, it is called a specific graph equivalence class or specific graph and is synonymous with a Markov equivalence class (Andersson, 1997). A specific graph, given data, produces a unique probability distribution. (This is the case for RA; as explained below, two different directed specific BN graphs can have the same distribution.) Since the number of graphs increases hyper-exponentially with the number of variables, searching the entire lattice for the best model can be very computationally expensive as the number of variables increases.

The colon notation for RA specific graphs represents marginal or conditional independence between variables. For example, G20 from Figure 2, having notation A:B:C:D, has the independence structure $(A \perp B, C, D), (B \perp C, D), (C \perp D)$; A, B, C, and D are all marginally independent of each other. G17, having notation AB:BC:D, has the independence structure $(D \perp A, B, C), (A \perp C \mid B)$; A, B, and C are all marginally independent of D, and A is conditionally independent of C given B. The overlap of B in the AB and BC relations encodes the conditional independence of A and C given B.

BN Lattice

The primary differences between RA and BN are two-fold: (1) BNs are directed and acyclic whereas RA graphs are undirected and can be cyclic or acyclic, and (2) some

BN graphs contain converging edges, called a V-structure, which encodes a unique independence structure not found in RA general graphs.

As in RA, there are BN general graphs and BN specific graphs, in the BN literature referred to as essential graphs and Markov equivalence classes respectively (Anderson, 1997). BN general graphs of Figure 3 represent unique independence structures of variables, edges, and edge orientations, where specific identities are not assigned to the variables. Figure 3 shows the hierarchy of BN general graphs for four variables. There are 20 BN general graphs in the lattice, i.e., 20 unique general independence structures for four variable BNs.

In Figure 3, general graphs are labeled BN1, BN2...BN19, BN20. Solid squares represent variables, edges are represented by directed arrows from one square to another. The dashed lines with arrows from one general graph to another represent the hierarchy of general graphs, with parent graphs being above child graphs. Child graphs result from the deletion of one edge from the parent graph. The bottom-right addition to the lattice that tabulates equivalent general graphs is discussed in the section titled "Additional representations of BN general graph equivalence classes."

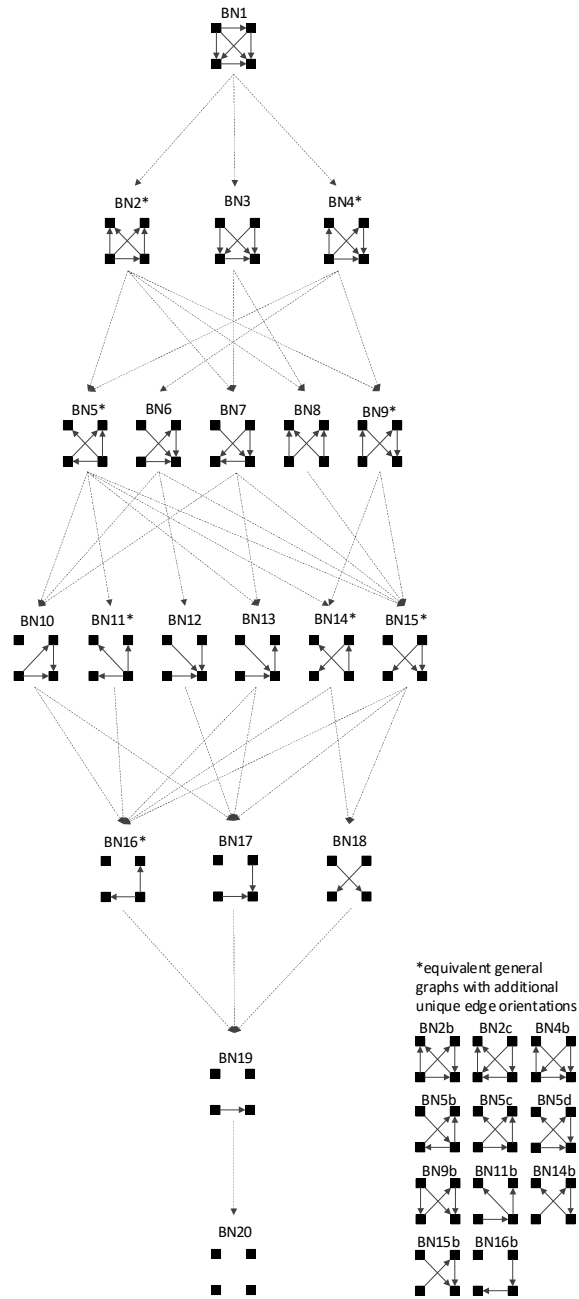


Figure 3 BN General Graph Lattice

The alphabetic notation that we use for BN specific graphs is derived from the RA notation described previously. For a BN graph without a V-structure, the notation is identical to the RA notation. As in RA, the colon represents marginal or conditional

independence among variables. For a BN graph with a V-structure, the notation adds subscripts to represent the independence relations encoded by the V-structure, which are unique to BNs and not found in RA. For example, BN17 from Figure 3, with labels A, B, C, D in order of top left, top right, bottom left, bottom right respectively, is given the notation $BCD_{B:C}:A$, where the colon between $BCD_{B:C}$ and A states the independency ($A \perp B, C, D$), namely that A is marginally independent of B, C, and D. The subscript $_{B:C}$ encodes marginal independence between B and C within the triadic BCD relation.

Generating the lattice of BN general and specific graphs

The BN literature on lattices predominately focuses on search algorithms to find the best BN given a scoring metric. Implicit in these search algorithms is a lattice of candidate graphs being explored in search of the best model. Chickering (2004) and others have shown the search problem to be NP-hard, with four variables there are 543 possible BNs, with 10 variables there are $O(10^{18})$ (Murphy, 1998). Because of this, the BN literature has focused less on the description of the exhaustive lattice of BN structures, and more on advancing search heuristics to efficiently identify the best BN given a scoring metric (Buntine, 1991a, 1991b; Cooper, 1992; Bouckaert, 1994; Heckerman, 1994; Chickering, 1995, 1997; Friedman, 1996, 2003; Larranaga, 1996; Koivisto, 2004; Malone 2011; and others).

Heckerman (1994) first showed that BNs with differing edge topologies can have the same independence structure and thus the same probability distribution, herein described as specific graphs. In contrast to heuristics that search all BNs, search

heuristics for BN specific graphs have proven to be more efficient because they reduce the dimensionality of search space (Chickering, 1995, 2002; Andersson, 1997; Gillispie, 2001; Studený, 2010; Tian, 2010; Zhang, 2012; Chen 2016; and others). For four variables, this reduces the search space from 543 BNs to 185 specific graphs (Anderson, 1997). Removing variable labels, these specific graphs can be summarized by 20 general graphs.

Building from the RA work of Klir (1986) and Zwick (2001), and the BN work of Pearl (1985, 1987, 1988, 2000), Verma (1990), Heckerman (1994), Chickering (1995), Andersson (1997), and others, the following procedure was used to generate the four variable BN general and specific graph lattice in a way that can be integrated with the four variable RA general graph lattice. While this procedure is applied in this paper to four variables, it could be used for any number of variables.

Procedure to generate the BN general and specific graph lattice for any number of variables

1. Generate all graphs for n variables by permuting all possible variable labels, edge connections and edge orientations. Eliminate any graphs with loops. The result is the set of all directed acyclic graphs for n variables.
2. For each directed acyclic graph, evaluate the specific independence structure following the d-separation procedure (MIT, 2015) to generate the exhaustive list of independencies for each graph.

3. Organize graphs with the same labeled independencies into specific graph equivalence classes.
4. From each specific graph equivalence class, select a single edge topology to represent the general graph equivalence class. List any additional equivalent general graphs with unique edge topologies separately, as done in Figure 3.
5. Organize general graphs into levels based upon the number of edges in each general graph and link nested general graphs in the lattice to reflect parent-child general graphs.

Additional representations of BN general graph equivalence classes

There are 20 general graphs in the four variable BN lattice. However eight of these, marked with asterisks in Figure 3, can be represented by additional unique edge topologies which, however, result in identical probability distributions when applied to data. These are BN2*, BN4*, BN5*, BN9*, BN11*, BN14*, BN15*, and BN16*. Thus, for example, BN2b has edge orientations that are different from (and cannot be mapped onto) those of BN2*, but when variables are labeled in BN2* and BN2b, identical probability distributions result. This property is unique to BN and is not found in RA, in which there is a single unique representation of each RA general graph.

Integrating the Rho, RA and BN Lattices

This section integrates the RA and BN general graph lattices using the four variable Rho lattice (Klir, 1985). The joint RA-BN lattice of general graphs presented in this paper expands the set of general graphs with unique independence structures

beyond what was previously available by either RA alone or BN alone. The lattice identifies general independence structures unique to RA, general independence structures unique to BNs, and general independence structures that are equivalent across RA and BNs. Where two or more RA or BN graphs have the same general independence structure regardless of variable labels, they are equivalent.

Lattice of four variable Rho graphs

The four variable Rho (ρ) lattice of Figure 4 (adapted from Klir, 1985, p. 237) is a simplification of the RA lattice of general graphs of Figure 2. The Rho lattice represents all possible undirected relations between four variables, an even more general lattice than that of the RA general graph lattice and general enough to map both RA and BN four variable general graphs to one of the eleven represented structures. Solid dots represent variables; lines connecting dots represent relations between variables. In terms of the RA general graph lattice

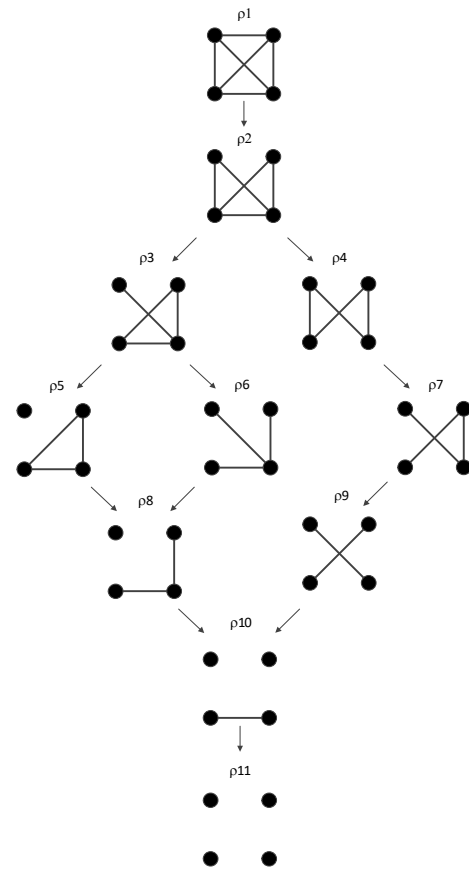


Figure 4 Rho Lattice

of Figure 2, if two variables are directly connected by any box (a relation of arbitrary ordinality), a line connects them in the Rho lattice. Arrows from one Rho graph to another represent hierarchy, i.e., the generation of a child graph from a parent graph.

The graph ρ_1 represents maximal connectedness, or interdependence, between variables, and the graph ρ_{11} represents independence among all variables. Graphs in-between ρ_1 and ρ_{11} represent a mix of dependence and independence among variables. Each RA or BN graph corresponds to one, and only one, of the eleven Rho graphs.

Equivalent RA and BN general graphs

Out of 20 RA general graphs and 20 BN general graphs, there are 10 RA general graphs, comprising all of the acyclic graphs in the RA lattice that are equivalent to BN general graphs. Each of these RA-BN equivalent pairs corresponds to one of the 11 Rho graphs from Figure 4, with the exception of ρ_4 . ρ_4 has corresponding RA and BN general graphs, but these do not have equivalent independence structures, and are discussed in the following section on non-equivalent RA and BN general graphs.

Figure 5 shows an example of equivalent RA and BN graphs, namely G7 and BN2*, respectively. Labeled variables in G7 results in independencies $(A \perp B \mid C, D)$ and thus the RA specific graph notation ACD:BCD. Assigning labels to BN2* yields the same independencies and thus the same specific graph ACD:BCD.

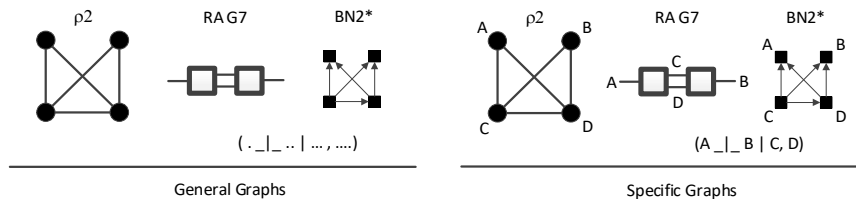


Figure 5 RA G7 and BN2* general and specific graph example

Table 9 shows the list of all equivalent Rho, RA and BN four variable general graphs, an example of their specific graph notation, and their independences. These specific graph examples align with the BN general graphs of Figure 3 assuming labeling of nodes A, B, C, D in the order of top left, top right, bottom left, bottom right.

Table 9 Rho, RA and BN equivalent graphs

| Rho graph | RA general graph | BN general graph | Specific Graph Example (RA notation) | Independencies |
|-------------|------------------|------------------|--------------------------------------|--|
| ρ_1 | G1 | BN1 | ABCD | no independencies |
| ρ_2 | G7 | BN2* | ACD:BCD | (A _ _ B C, D) |
| ρ_3 | G10 | BN5* | BCD:AD | (A _ _ B, C D) |
| ρ_5 | G13 | BN10 | BCD:A | (A _ _ B, C, D) |
| ρ_6 | G15 | BN11* | AD:BD:CD | (A _ _ B, C D), (B _ _ C D) |
| ρ_7 | G16 | BN14* | AD:BC:BD | (A _ _ B D), (C _ _ A, D B) |
| ρ_8 | G17 | BN16* | BD:CD:A | (B _ _ C D), (A _ _ B, C, D) |
| ρ_9 | G18 | BN18 | AD:BC | (A, D _ _ B, C) |
| ρ_{10} | G19 | BN19 | CD:A:B | (B _ _ C, D), (A _ _ B, C, D) |
| ρ_{11} | G20 | BN20 | A:B:C:D | (A _ _ B, C, D), (B _ _ C, D), (C _ _ D) |

Non-equivalent RA and BN general graphs

In addition to the 10 equivalent RA and BN general graphs, there are 10 general graphs unique to the RA lattice and 10 general graphs unique to the BN lattice. All 10 non-equivalent RA general graphs in the four variable lattice are cyclic and require iteration to generate their probability distributions. BNs are acyclic and have analytic solutions, so there are no BN graphs that are equivalent to these cyclic RA graphs. Since cyclic RA graphs are undirected, one might think that there could be some equivalent directed acyclic BN graphs, but this is not the case, because BN graphs that are acyclic when directions are considered but cyclic if directions are ignored have V-structure

interpretations, as described previously. All 10 non-equivalent BN graphs have such V-structures, which encode independence relations unique to BNs.

To illustrate: the structure $A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A$ is cyclic and not a legitimate BN structure, but the directed structure of $A \rightarrow B, B \rightarrow C, C \rightarrow D, A \rightarrow D$, which has the same undirected links, is not cyclic, and is a legitimate BN structure, namely BN9b. However, this latter structure is not interpreted as a set of dyadic relations, which would be written in RA notation as $AB:BC:CD:AD$. Rather, the V-structure consisting of $C \rightarrow D$ and $A \rightarrow D$ is interpreted as a triadic relation, which contributes a $p(D|AC)$ to the probability expression, $p(A) p(B|A)p(C|B) p(D|AC)$, which does not correspond to any RA structure.

Joint lattice of RA and BN general graphs

The joint lattice of RA and BN general graphs is organized using the Rho lattice (Klir, 1985) of Figure 4, the RA general lattice of Figure 2 (Klir 1985, Krippendorff 1986) and the BN general lattice of Figure 3. All 20 RA general graphs and all 20 BN general graphs for each Rho graph are shown in the joint lattice of RA and BN general graphs of Figure 6.

In Figure 6, within each Rho graph, where RA and BN general graphs are equivalent, that is, their independence structures are identical, the BN graph is placed underneath the RA equivalent graph. Where RA or BN graphs are not equivalent, representing an independence structure unique to RA or to BN, they stand alone side-

by-side. Arrows from one graph to another in the joint lattice represent the hierarchy of the RA lattice only.

Conclusions

The joint lattice of RA and BN general graphs for four variables increases the number of general graphs with unique independence structures from 20 in the four variable RA lattice and 20 in the four variable BN lattice to 30 in the joint RA-BN lattice, and when variable labels are added, increases the number of unique specific graphs from 114 in the RA lattice and 185 in the BN lattice to 238 in the joint lattice.

The integration of the two lattices offers a richer and more expansive way to model complex systems leveraging the V-structure unique to BN graphs and the allowability of cycles in RA graphs. The joint RA-BN lattice of general graphs presented in this paper expands the set of general graphs with unique independence structures (or, equivalently, with unique interdependence structures) beyond what was previously available by either RA alone or BN alone, thus allowing for representations of complex systems which are (i) more accurate relative to data and/or (ii) simpler and thus more comprehensible and more generalizable than would be possible by modeling only with RA or only with BN. This joint lattice thus demonstrates how these two related frameworks – RA and BN – both members of the family of probabilistic graphical modeling methodologies, can be integrated into a unified framework. Extension of this work will include designing algorithms to search this joint RA-BN lattice, analysis of RA

and BN predictive models in which the IV-DV distinction is made, consideration of “hybrid” RA-BN models, and other topics.

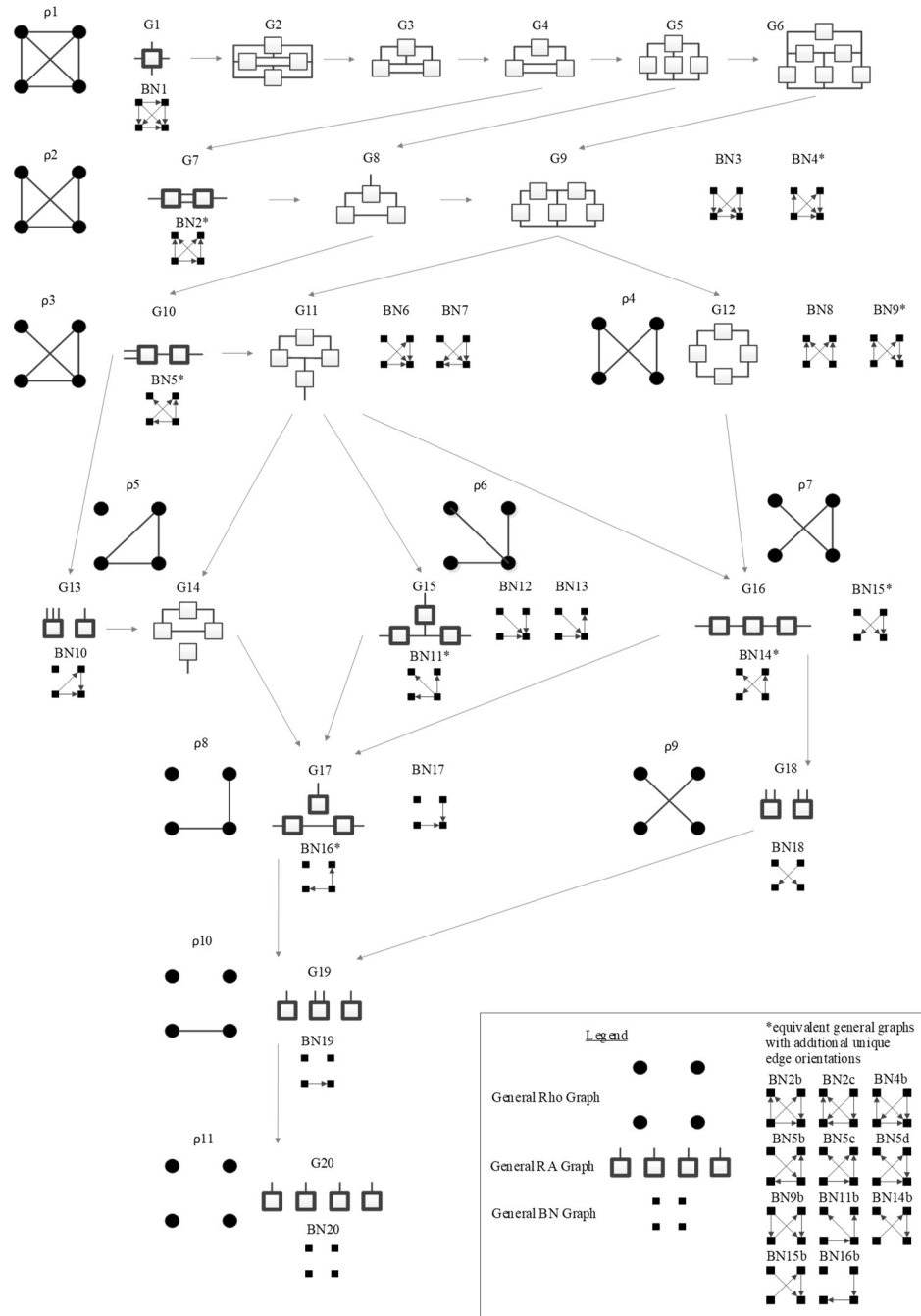


Figure 6 Joint Lattice of RA and BN General Graphs

Chapter 2

Paper 2 - Graphical Models in Reconstructability Analysis and Bayesian Networks

Harris, M.; Zwick, M. Graphical Models in Reconstructability Analysis and Bayesian Networks. *Entropy* 2021, 23, 986. <https://doi.org/10.3390/e23080986>

Co-authored by Marcus Harris and Martin Zwick.

Author contribution statement: Marcus Harris performed the conceptualization, formal analysis, writing original draft, and visualization. Dr. Martin Zwick performed conceptualization, writing review and editing, and supervision.

DOI: <https://doi.org/10.3390/e23080986>

Link to published paper: <https://archives.pdx.edu/ds/psu/36135>

Abstract

Reconstructability Analysis (RA) and Bayesian Networks (BN) are both probabilistic graphical modeling methodologies used in machine learning and artificial intelligence. There are RA models that are statistically equivalent to BN models and there are also models unique to RA and models unique to BN. The primary goal of this paper is to unify these two methodologies via a lattice of structures that offers an expanded set of models to represent complex systems more accurately or more simply. The conceptualization of this lattice also offers a framework for additional innovations beyond what is presented here. Specifically, this paper integrates RA and BN by developing and visualizing: (1) a BN neutral system lattice of general and specific graphs, (2) a joint RA-BN neutral system lattice of general and specific graphs, (3) an augmented RA directed system lattice of prediction graphs, and (4) a BN directed system lattice of prediction graphs. Additionally, it (5) extends RA notation to encompass BN graphs and (6) offers an algorithm to search the joint RA-BN neutral system lattice to find the best representation of system structure from underlying system variables. All lattices shown in this paper are for four variables, but the theory and methodology presented in this paper are general and apply to any number of variables. These methodological innovations are contributions to machine learning and artificial intelligence and more generally to complex systems analysis. The paper also reviews some relevant prior work of others so that the innovations offered here can be understood in a self-contained way within the context of this paper.

Introduction

Reconstructability Analysis (RA) and Bayesian Networks (BN) are both probabilistic graphical modeling methodologies. A probabilistic graphical model uses a graph (or hypergraph) to encode independencies and dependencies between variables and probability theory to encode the precise nature of the relations between variables. Graphs are either undirected or directed. RA graphs include undirected graphs (or hypergraphs) that have loops or do not have loops. BN graphs are directed graphs that do not have cycles. (“Loops” here refer to undirected graphs; “cycles” refer to directed graphs.) RA and BN graphs can represent independence structures that are unique to each methodology, and also independence structures that are the same in both methodologies. For RA models without loops and for all BN models, variable independencies can be represented in closed algebraic (factorized) form. For RA models with loops, solutions require iterative calculations. The value of integrating these two methodologies lies in the fact that the RA lattice of structures offers potential models of complex systems not found in BNs, while BNs are a more widely used analytical approach than RA and also include unique models. Combining the candidate models of the two methodologies thus offers a more expressive framework than either alone. It also does so in an organized and coherent way that allows for future possible extensions discussed in the section titled “Discussion”.

RA is a data modeling approach developed in the systems community (Ashby, 1994; Broekstra, 1979; Cavallo, 1979; Conant, 1981, 1988; Klir, 1976, 1985, 1986;

Krippendorff, 1979, 1981, 1986; Willet, 2004; Zwick, 2001, 2004a, 2004b, 2010, 2018) that combines graph theory and information theory. Its applications are diverse, including time-series analysis, classification, decomposition, compression, pattern recognition, prediction, control, and decision analysis (Zwick, 2001). It is designed especially for nominal variables, but continuous variables can be accommodated if their values are discretized. RA could in theory accommodate continuous variables; however, this extension of the methodology has yet to be formalized. Graph theory specifies the structure of the model: if the relations between the variables are all dyadic (pairwise), the structure is a graph; if some relations have higher ordinality, the structure is a hypergraph. In speaking of RA, the word 'graph' will henceforth include the possibility that the structure is a hypergraph. The structure is independent of the data except for specification of variable cardinalities. In RA, information theory uses the data to characterize the precise nature and the strength of the relations. Data applied to a graph structure yields a probabilistic graphical model of the data.

RA has three primary types of models: variable-based models without loops, variable-based models with loops and state-based models (where individual states of variables specify model constraints) that nearly always have loops. Models that do not have loops have closed-form algebraic solutions; those that have loops require iterative proportional fitting. In RA, graphs are undirected, although directions are implicit if one variable is designated as the response variable (dependent variable or DV), while all other variables are designated as explanatory variables (independent variables or IVs).

In principle, there could be more than one DV, but in the discussion that follows, a single DV is assumed. If the IV-DV distinction is made, the system is 'directed' and the primary aim is prediction of the DV given the IVs; if no IV-DV distinction is made, the system is 'neutral' and the primary aim is to characterize the nature of relations among all variables.

RA models are undirected graphs that either have or do not have loops, where a 'loop' is the presence of circularity in a set of undirected links. We reserve the word 'cycle' and 'acyclic' for circularity or lack thereof in directed graphs, which are used in BN and not in RA. An undirected graph having a loop can become an acyclic graph for certain assignments of link directions. For example, an RA model that posits relations between A and B, between B and C, and between A and C has a loop, but if directions are assigned in a BN model so that these relations are $A \rightarrow B$, $B \rightarrow C$, and $A \rightarrow C$, the resulting graph is acyclic.

Graphs are general or specific. A general graph identifies relations among variables that are unlabeled, i.e., variables whose identity is not specified; a specific graph labels (identifies) the variables. For example, for a system consisting of variables A, B, and C, AB:BC is a specific graph where nodes A and B are linked and B and C are also linked. Specific graphs AB:BC, BA:AC and AC:CB are all instances of the same general graph that has a unique independence structure regardless of variable labels. In this notation, the order of variables in any relation is arbitrary, as is the order of the relations. For example, CB:BA is identical to AB:BC. Relations include all of their

embedded relations. For example, ABC includes embedded relations AB, AC and BC and the univariate margins A, B, and C.

The lattice of graphs for a neutral or a directed system with or without loops depends upon the number of variables in the data. For a three-variable neutral system allowing loops there are five general graphs and nine specific graphs; for four variables there are 20 general graphs and 114 specific graphs. The number of graphs increases hyper-exponentially with the number of variables. In the confirmatory mode, RA can test the significance of a single model—a hypothesis being tested—relative to another model used as a reference. In the exploratory mode, RA can search the lattice of graphs for models that are statistically significant and best represent the data with maximal information captured and minimal complexity.

Bayesian Networks (BN) are another probabilistic graphical modeling approach to data modeling that is closely related to RA. Indeed, where BN overlaps RA the two methods are equivalent, but with respect to neutral systems, RA and BN each has distinctive features absent in the other methodology. For directed systems; however, where prediction of a single dependent variable is the aim, RA encompasses all models found in BN under the convention used in this paper that all nodes except for parent nodes within a V-structure are allowed to be the variable being predicted; this inclusion of the BN directed system lattice within the RA lattice will be shown later in this paper.

BNs have origins in the type of path model described by Wright (1921, 1934), but it was not until the 1980s that BNs became more formally established (Neapolitan, 1989; Pearl, 1985, 1987, 1988). As does RA, BN combines graph theory and probability theory: graph theory provides the structure and probability theory characterizes the nature of relationships between variables. BNs are represented by a single type of graph structure; a directed acyclic graph, which is a subset of chain graphs, also known as block recursive models (Lauritzen, 1996). BNs can be represented more generally by partially directed acyclic graphs (PDAG), a subset of chain graphs where edge directions are removed when directionality has no effect on the underlying independence structure. Discrete variables are most common in BNs, but BNs accommodate continuous variables without discretization (Driver, 1995). In principle RA could also accommodate continuous variables but this feature has not yet been implemented. For a three variable BN lattice, there are 5 general graphs and 11 specific graphs; for four variables there are 20 general graphs and 185 specific graphs with unique probability distributions. In the confirmatory mode, BNs can test the significance of a model relative to another model used as a reference (Tang, 2012); in the exploratory mode, BNs can search for the best possible model given a scoring metric. BNs are used to model expert knowledge about uncertainty and causality (Neapolitan, 1989; Pearl, 1985) and are also used for exploratory data analysis with no use of expert knowledge (Rebane, 1987). Like RA, BN applications in machine learning and artificial intelligence are broad including

classification, prediction, compression, pattern recognition, image processing, time-series, decision analysis and many others.

The joint RA-BN lattice of neutral system general and specific graphs and the accompanying search algorithm developed in this paper expands both RA and BN beyond what was previously available by either RA alone or BN alone, thus providing a more complete ensemble of models for the representation of complex systems. When prediction of a single dependent variable (DV) is the aim, the RA directed system lattice encompasses the BN directed system lattice under the strict convention used in this paper that excludes a parent node with a V-structure being the DV. However, we also show that when this constraint is relaxed so that a parent node within a V-structure can be the DV, BN models can offer predictions unique to BN. We also show that (under the above convention) the BN directed system lattice reduces the size of the full BN neutral system lattice by retaining only graphs that give unique predictions of the DV, significantly reducing the search space to find the best BN when prediction of a single DV is the aim. Finally, this paper develops an augmented RA directed system lattice which expands the conventional RA lattice of prediction graphs to include naïve Bayes equivalent graphs. This augmented lattice encompasses graphs in the BN directed system lattice and allows for models of complex systems which are (i) more predictive and/or (ii) simpler and thus both more comprehensible and more generalizable than models restricted to the conventional RA directed system lattice.

RA Lattice

RA Neutral Systems

All lattices shown in this paper are for four variables, but the theory and methodology presented in this paper are general and apply to any number of variables. RA neutral systems include only independent variables, i.e., there is no concept in such systems of a dependent variable. A neutral system model thus represents the relationships, graphically and probabilistically, between all the (independent) variables. The graphical representation specifies the independencies among variables. When data are then applied, probabilities represent the strength of the relationships between dependent variables. Neutral system graphs are commonly used in applications where variable clustering is important, such as computer vision and social and biological network analyses. Neutral system analysis is more computationally demanding than directed system analysis, so when one is really interested in predicting specific variables, directed system models are more convenient.

The four-variable RA lattice of neutral system general graphs (Figure 7), (Klir, 1985; Krippendorff, 1979), represents all four-variable RA graphs with unique independence structures. Bold graphs do not have loops while non-bold graphs have loops. In these graphs, lines (including branching lines) are variables; boxes are relations. Where only two lines extend from a box, the relation is dyadic. If more than two lines extend from a box, the graph is a hypergraph. Where two or more specific graphs have the same independence structure, regardless of variable labels, they are

part of the same general graph equivalence class. For example, the left-most and right-most variables in G7 are independent of one another given the two central variables that connect both relations; this results in the general independence structure $(. \perp .. | \dots, \dots)$, where each different number of dots indicates a different variable, but does not specify its actual identity. The expression says that the first variable is independent (“ \perp ” is the symbol used in this paper for independence) of the second variable given (“ $|$ ” is the symbol used in this paper for “given”) the third and (the comma “,” represents a logical “and”) fourth variables.

G1 is the most complex general graph, in which the variables are connected in a tetradic relation. Graphs below G1 reflect increasingly less complex decompositions of G1, ending with G20 which has complete independence among the variables. Arrows from one general graph to another represent hierarchy such that going from the parent graph (the source of the arrow) to the child graph (the terminus of the arrow) results from deleting one relation from the parent graph.

In this paper, when the variables of a general graph are labeled in RA or BN, it is called a specific graph, which is a unique probabilistic model given the data. For RA, given data and after labeling all the variables, there is only one specific graph for any general graph. By contrast, as explained in the section titled “BN Lattice,” (beginning in the sub-section titled “BN neutral systems: Lattice of BN general graphs”), two or more topologically different BN general graphs can have the same probability distribution; such equivalent graphs have the same underlying set of independencies even though

they are topologically different; they are said to constitute a ‘Markov equivalence class’
 (Verma, 1990).

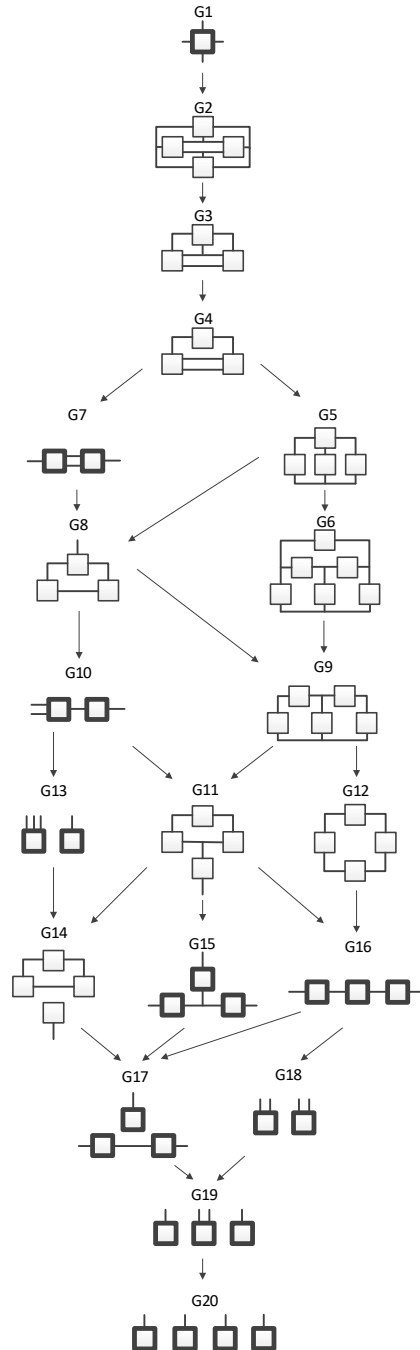


Figure 7 Lattice of four-variable RA neutral system general graphs.

RA graphs can include pairwise and non-pairwise relations. For example, graph G15 has four lines (variables) and three boxes (relations). One line connects to all three boxes, meaning one variable is included in all three relations, and separately a single line representing one of the other three variables extends from each box. Because only two lines extend from any given box, all relations in G15 are pairwise (dyadic). Figure 8 shows G15 with labels (A, B, C, D) added for the variables, yielding a specific structure having dyadic relations AD, BD, and CD. In RA notation, this graph is AD:BD:CD, the colon represents independence among relations. The notation AD:BD:CD encodes the independencies $(A \perp B, C \mid D)$, $(B \perp C \mid D)$. The example in Figure 8 represents one of four specific graphs for the general graph G15, the other possible permutations are AB:AC:AD, BA:BC:BD, CA:CB:CD. These permutations have the same general independence structure $(. \perp ., \dots \mid \dots)$, $(\dots \perp \dots \mid \dots)$, but given data, produce different conditional probability distributions.

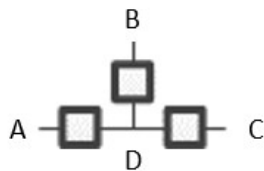


Figure 8 RA specific graph G15, AD:BD:CD

In contrast to graph G15 which includes dyadic relations only, graph G13 in Figure 9 is a hypergraph, with three lines extending from one box and a single line extending from the other. This could, for example, represent four variables A, B, C and D, where A, B, and C label the three lines extending from one box, and D labels the

single line extending from the other box. Figure 9 shows this specific graph, which in RA notation is ABC:D with the independence structure of $(D \perp A, B, C)$. This example represents one of four specific graphs for general graph G13, the other three being, ABD:C, ACD:B, and BCD:A with independencies of $(C \perp A, B, D)$, $(B \perp A, C, D)$, and $(A \perp B, C, D)$, respectively. Given data, each of these four specific graphs (ABC:D, ABD:C, ACD:B, and BCD:A) generates a unique probability distribution.

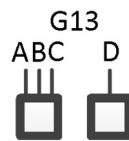


Figure 9 RA specific graph G13

Figure 10 shows all of the general graphs from Figure 7 as well as all of the specific graphs associated with each general graph. There are 20 general graphs in the RA lattice and 114 specific graphs.

Searching the RA Neutral System Lattice

The data are the top of the lattice, i.e., G1 ABCD, and one searches the lattice to find a good representation of the data. The lattice can be searched from the top down or from the bottom up or from some other starting model. Typically, a reference model, a specific graph, is selected to begin the search. Commonly, it is the independence (bottom) model, G20 A:B:C:D from Figure 10, that is selected as the reference model, and the lattice is searched upward to find the best model. The lattice may also be searched downward starting from the saturated (top) model, G1, or from a reference

model in-between the bottom or top, searching up or down. The starting model does not have to be the reference model, but this is often the case.

Commonly, when the lattice is being searched, the goal is to find a model (a specific graph) that adequately represents but is less complex than the data. This best characterizes a search downwards that (typically) starts from G1. When searching down the lattice, the goal is to search as far down the lattice as possible, resulting in the greatest complexity reduction from the reference model, while incurring the least amount of information loss, so that the model still adequately represents the data. Finding a simpler representation of the data reduces the complexity of the system under observation, allowing for greater understanding of the most important underlying relations. Alternatively, the goal is to find a model, a specific graph, that captures as much of the information in the data as possible, as long as its difference from mutual independence of the variables, i.e., G20 in Figure 10, is defensible, so the model is not overfit, and its application to new data is likely to be more successful. This best characterizes a search upwards that (typically) starts from G20. For directed systems where prediction of a single DV is the aim, a high information model is one that gives maximal reduction of the Shannon entropy (uncertainty) of the DV.

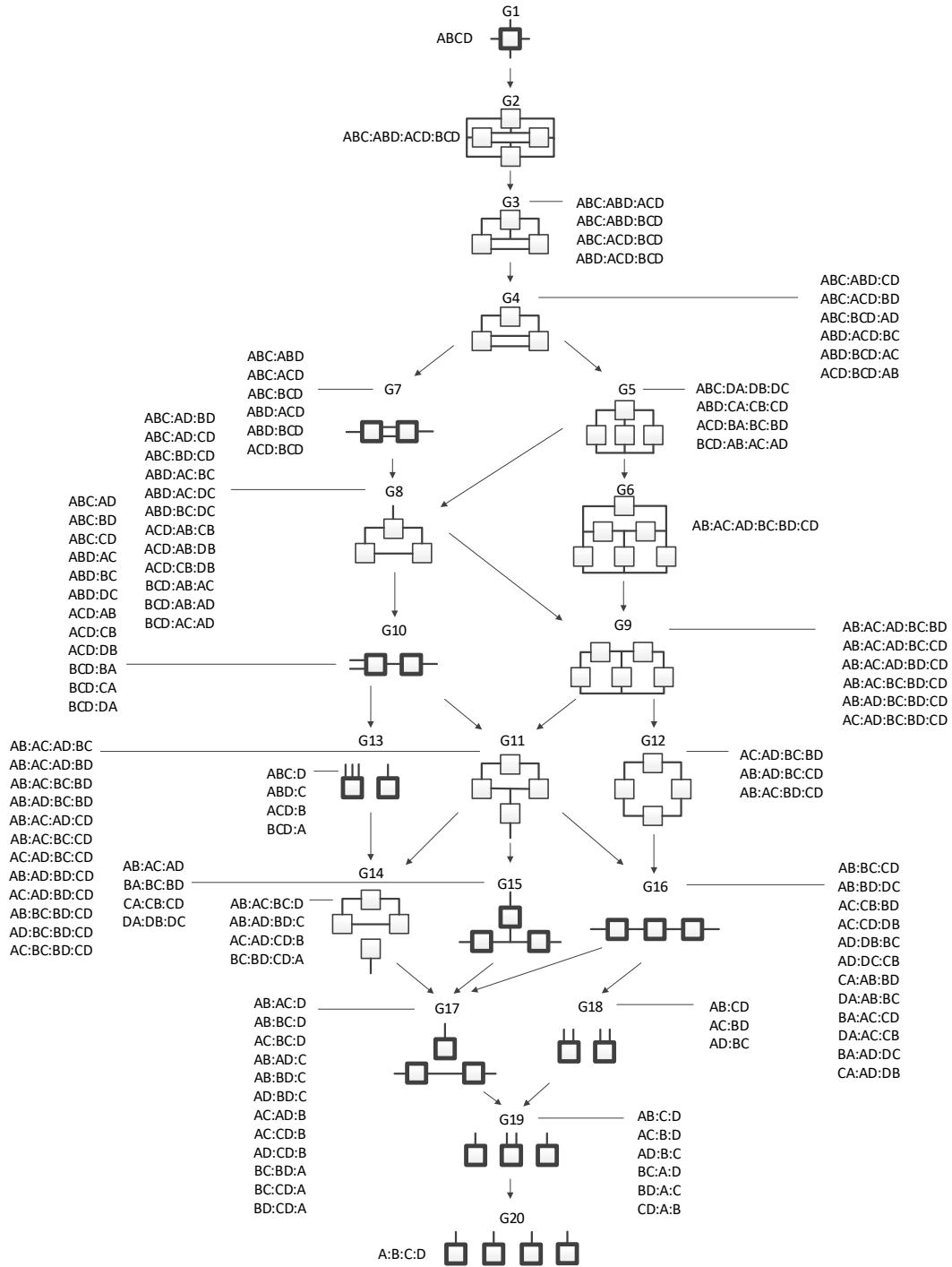


Figure 10 Lattice of RA neutral system general and specific graphs

Given data, specific graphs can be tested for statistical significance. The Chi square statistical test can be used to test the difference between any candidate model

and a reference model, usually the data, G_1 , or the independence model, G_{20} . As an alternative or in addition to such a statistical test, the Bayesian Information Criteria (BIC) and the Akaike Information Criteria (AIC) are among the other measures that can be used to decide on the best model.

RA Directed Systems

Conventional Directed System Lattice

The RA lattice of directed systems shown in Figure 11 is a sub-lattice of the complete neutral system lattice of Figure 7. The purpose of the directed system lattice is to organize models that make an IV-DV (explanatory-response) distinction and where prediction of the DV is the sole aim. In contrast, the neutral system lattice organizes models that do not make any IV-DV distinction; these models do not focus on a single response variable. There are fewer general graphs in the directed system lattice compared to the neutral system lattice because, by convention, we care only about models whose predictions of the DV are different and are not interested in identifying relations among the IVs. The word 'directed' in RA 'directed systems' has a meaning that is different from the meaning of the same word in BN 'directed acyclic graphs'. In RA 'directed systems', this word means that the focus of modeling is on the relation of the dependent variable to the independent variables. It does not imply directionality of edges from the IVs to the DV as this word means in BN 'directed acyclic graphs'.

In the neutral system lattice of Figure 7, any of the variables can be part of any relation. In contrast, in the standard directed system lattice, by convention, all of the IVs

are always included in one of the relations (the “IV relation”); the other relations in the model include predictive IV-DV interactions (or the DV alone if there are no such interactions). In this paper, the DV in directed system specific graphs is called “Z” and the IVs are called A, B, C, and so on. For example, the first specific graph listed under G3 from Figure 11, ABC:ABZ:ACZ, has all three IVs in the first relation, followed by two IV-DV relations. Aside from allowing for the presence of relations among the IVs (without specifying any such relations), the model says that there is a relation in which A and B might predict Z and another relation in which A and C might predict Z; the net predictive relation between A, B, C and Z is a maximum entropy fusion of these two predictive relations.

General graph G13 (ABC:Z) from Figure 11 represents independence between the IVs (ABC) and the DV (Z), thus there is no relation between the IVs and the DV, and graph G1 (ABCZ, which is not written as ABC:ABCZ because ABC is embedded in ABCZ) represents complete dependence among the IVs and the DV. It should be noted; however, that the directed system lattice of Figure 11 is not entirely exhaustive. What restricts this lattice is that all models include the “IV relation”; this makes these models hierarchically nested, and amenable to standard statistical tests. There are additional predictive graphs where this restriction is dropped that produce different predictions of Z than the models of Figure 11; these additional graphs are discussed in the following section titled “Augmented directed system lattice”.

Figure 11 shows all directed system general and specific graphs for four variables. The graphs that are greyed represent graphs from the neutral system lattice from Figure 11 that are not part of the directed system lattice because they do not offer unique predictions of the DV. There are nine directed system general graphs and 19 specific graphs in contrast to the neutral system lattice, which has 20 general graphs and 114 specific graphs.

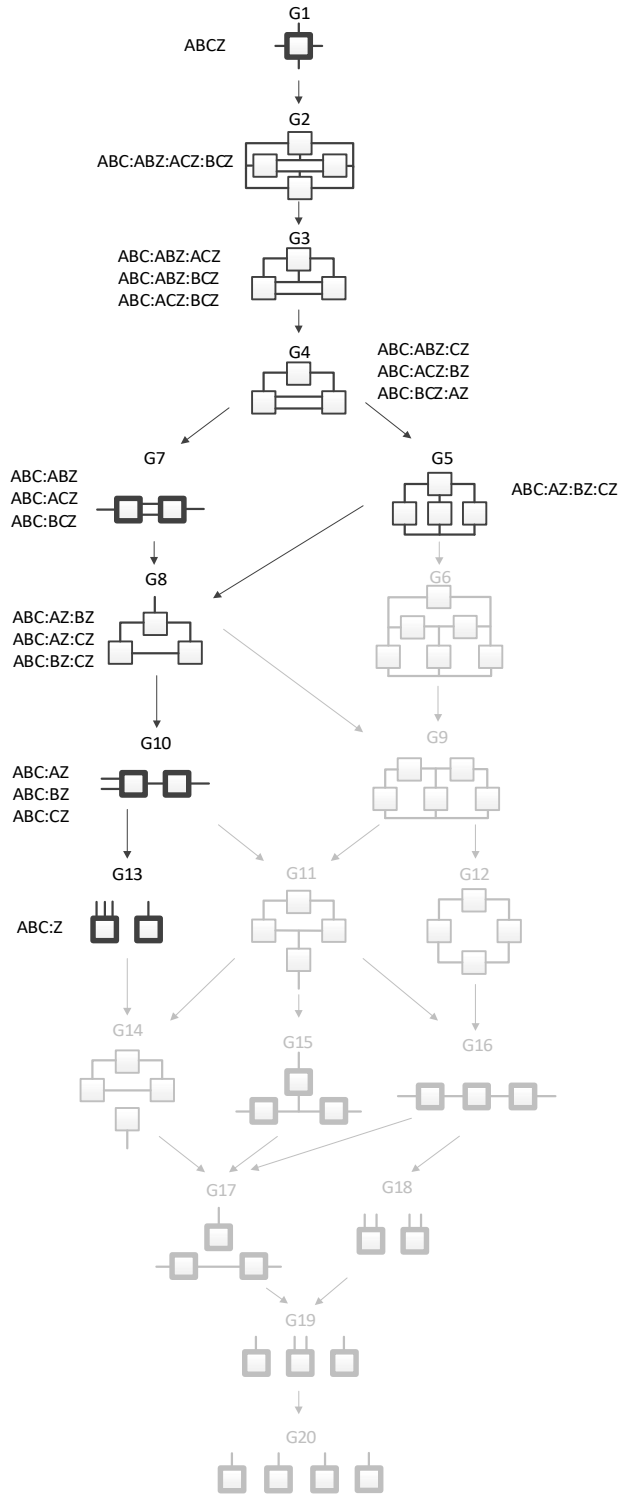


Figure 11 Conventional RA directed system lattice

Augmented Directed System Lattice

Figure 12 augments the conventional directed system lattice (on the left) of Figure 11 with a lattice of additional predictive graphs (on the right). These additional graphs offer unique analytical results, but that are not typically included when searching the hierarchically restricted directed system lattice.

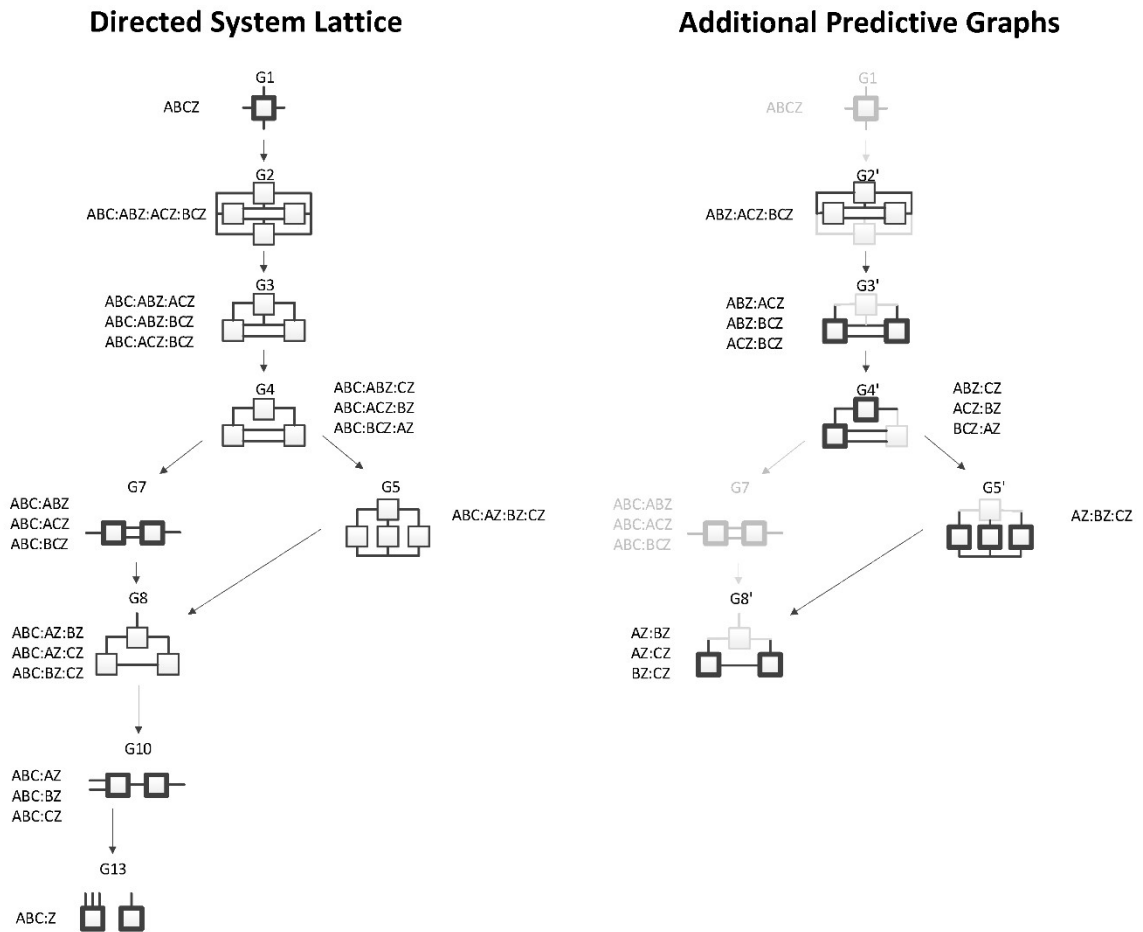


Figure 12 Conventional RA directed system lattice and additional predictive specific graphs

In Figure 12, the graphs in the Additional Predictive Graphs lattice are denoted by an apostrophe to identify that the original graph was altered by removing the IV

relation. For example, the bottom relation in graph G2, interpreted as the IV relation, ABC, was removed to produce an additional predictive graph G2'. This general graph has only one specific graph, ABZ:ACZ:BCZ, which is analytically different from G2 (ABC:ABZ:ACZ:BCZ) because the ABC term in graph G2 imposes a constraint among the IVs that is not imposed in graph G2'. Because G2' does not follow the standard directed system convention of including the IV terms in a first relation, it produces a different prediction of Z. The apostrophe-marked graphs are less complex than the graphs from which they are derived, and so should also be considered in searches for good predictive models. G5' and G8' from Figure 12 represent naïve Bayes equivalent RA graphs; G4' is also a naïve Bayes-like graph. This is discussed in the section titled "BN directed systems".

A merger of the conventional directed system lattice with the additional predictive graphs of Figure 12 gives the augmented directed system lattice in Figure 13. The specific graphs from G2' and G3' from Figure 12 are members of general graphs G3 and G7, respectively. Three general graphs are added to the augmented lattice, namely G10, G15 and G17; these are G4', G5', and G8' from Figure 12, the naïve Bayes or naïve Bayes-like equivalent RA general graphs. All of the specific structures that are added to the augmented lattice are denoted in bold letters in Figure 13. G13 is the independence model for the conventional directed system graphs. The augmented lattice also includes G20 A:B:C:D, which is the natural independence model for the additional predictive graphs that do not include the IV term (ABC). Including these additional predictive

graphs in the directed system lattice increases the number of predictive general graphs from nine in the conventional directed system lattice to 12 in the augmented lattice and 19 specific graphs in the conventional lattice to 31 in the augmented lattice.

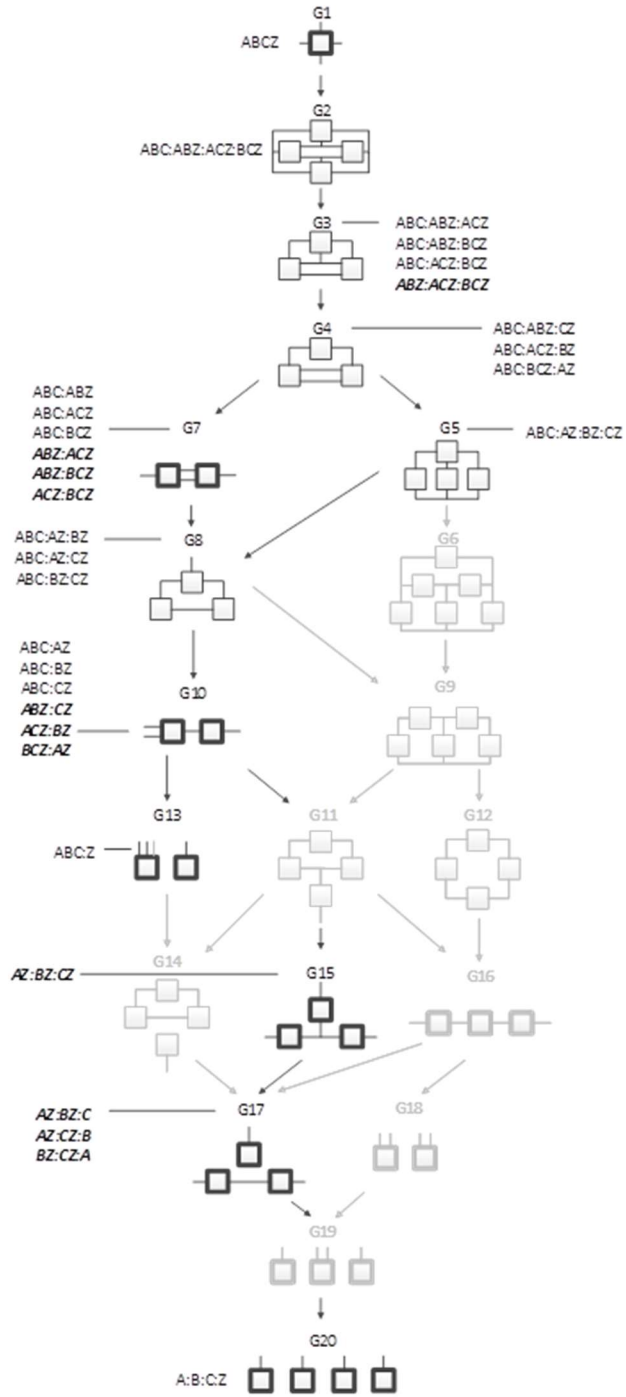


Figure 13 Augmented RA directed system lattice

BN Lattice

BN Introduction

A Bayesian Network model, like an RA model, is a type of probabilistic graphical model. BN modeling originated from path models in the early 1900s (Wright, 1921, 1934) and was expanded as a field of study in the late 1900s by Pearl (1985), Neapolitan (1989) and others.

BNs are directed graphs: nodes represent variables, and edges represent relations. The graph structure or topology (variables, edges, orientations of edges) encodes independencies, and thus also dependencies, among the variables identified in a particular graph. Since BNs are directed graphs, edges typically have arrows or some form of notation representing directionality: $A \rightarrow B$ means that variable B is dependent upon variable A. (This dependency might be interpreted as a causal influence of A on B, but in this paper, we will not address such causal interpretations of BNs.) A is the 'parent' of B, which means that they are dependent. One variable is independent of all other variables given its parents. For example, in the BN $A \rightarrow B \rightarrow C$, variable C is independent of A given B, since B is the parent of C.

A BN graph provides the structure from which a probability expression can be derived that describes the relation between variables. For example, the graph $A \rightarrow B$ provides the structure identifying the dependence between A and B, and probability values define the nature and strength of the relation between A and B. A unique feature of BNs versus other graphical models is in the independencies that are encoded when

two edges converge. For example, in $A \rightarrow B \leftarrow C$ the edges converge on variable B. If A and C are not directly connected by an edge, this convergence is called a V-structure (Chickering, 2002). This V-structure is interpreted as yielding the conditional distribution $p(B|A,C)p(A)p(C)$, which encodes dependence among A, B, and C, but marginal independence between A and C. The interpretation is that together, but being independent of one another, A and C influence or cause or allow one to predict B.

BNs are also acyclic graphs, meaning they have no closed paths following the arrows. For example, graph $A \rightarrow B \rightarrow C \rightarrow A$ is disallowed because it contains a cycle. Because BNs are acyclic, inference on all BN graphs can be performed in closed algebraic form.

The primary differences between RA and BN are two-fold: (1) BNs are directed and acyclic whereas RA graphs are undirected and can have loops or not have loops and (2) some BN graphs contain converging edges, that is one or more V-structures that encode unique independence relations not found in RA graphs. The absence of a V-structure in a BN graph results in this graph being equivalent to some (loopless) RA graph. The presence of a V-structure results in the graph not having an RA equivalent and thus being unique to BN. This is discussed below in the section titled "BN Neutral System General and Specific Graph Procedure" in connection with Table 12.

BN Neutral Systems

Lattice of BN General Graphs

As in RA, there are general BN graphs and specific BN graphs; in the BN literature general graphs are referred to as maximally oriented graphs (Meek, 1995), essential graphs (Andersson, 1997), equivalence classes of directed acyclic graphs (Gillispie, 2001), and partially directed graphs (PDAG) (Chickering, 2002).

In BN general graphs, the graph structure (variables, edges, and orientations of edges) results in a unique independence structure, where specific identities are not assigned to the variables. Figure 14, developed by Harris and Zwick (2021), shows all BN general graphs with four variables and their hierarchy. There are 20 BN general graphs in the lattice, i.e., 20 unique independence structures. The procedure to generate this lattice is outlined in the section titled “BN Neutral System General and Specific Graphs Procedure”.

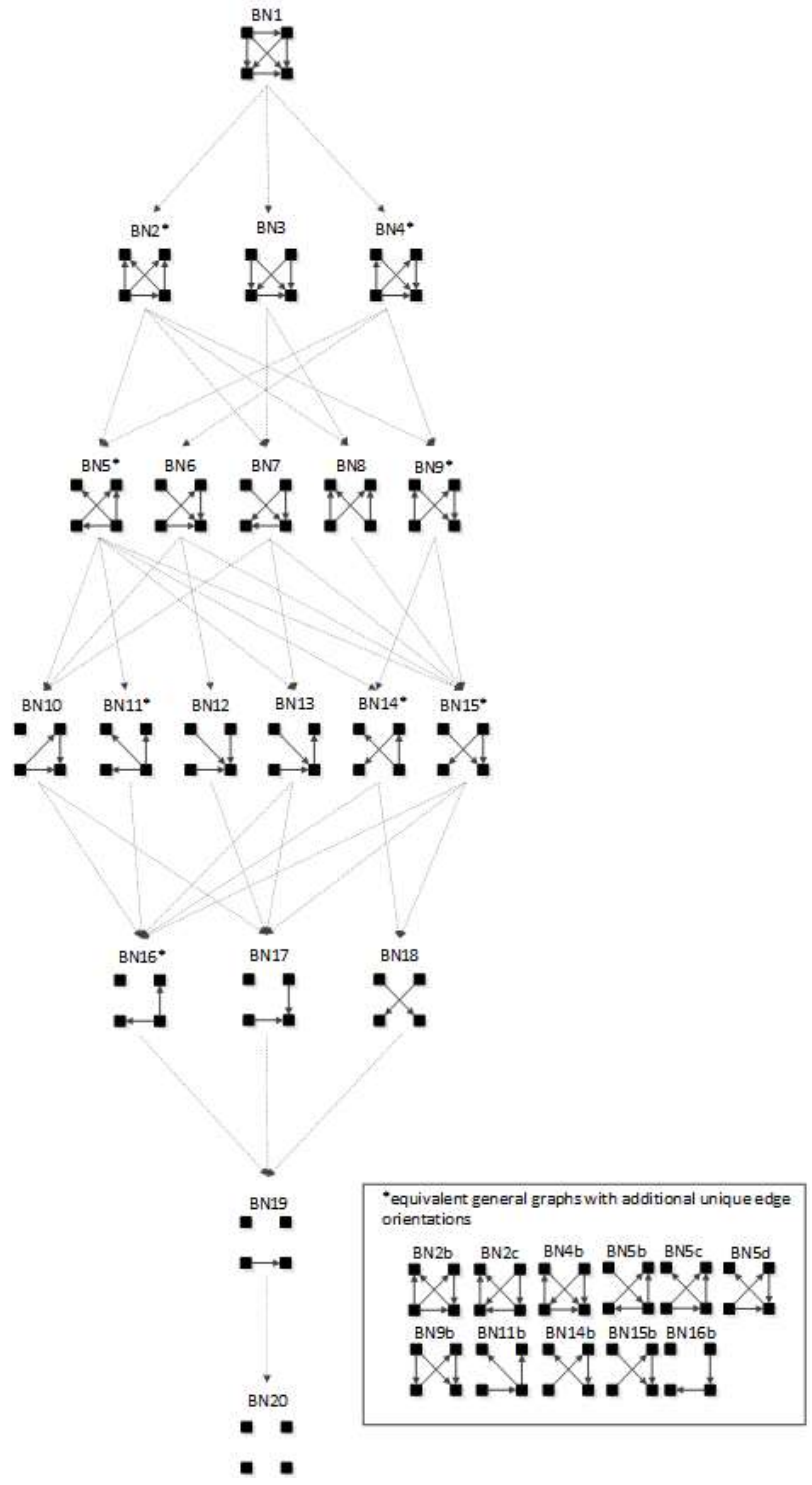


Figure 14 Lattice of BN neutral system general graphs

In Figure 14, general graphs are labeled BN1, BN2...BN19, BN20. Solid squares represent variables; edges are represented by directed arrows from one square to another, representing a parent–child dependency relationship. The dashed lines with arrows from one general graph to another represent the hierarchy of general graphs, with parent graphs being above child graphs. Child graphs result from the deletion of one edge from the parent graph⁷. The insert on the bottom right indicates structures that are topologically different from graphs in the lattice marked with asterisks but have identical independence structures to these marked graphs and thus are Markov equivalent (the topological difference cannot be removed by any labeling of the variables). For example, BN2b and BN2c in the insert are topologically different but have the same independence structure as BN2* in the lattice. These additional representations are discussed below in the section titled “Additional representations of BN general graphs”.

⁷ Not included in the published paper: unlabeled graphs in the four variable BN lattice are representative of equivalence classes, i.e. unique unlabeled independence structures. A “child” general graph in the lattice is generated by the deletion of one edge from any member of a parent equivalence class. For example, deletion of an edge from BN2*, BN2b, or BN2C, all members of the same equivalence class, results in child graphs BN5*, BN7, BN8, and BN9*.

Table 10 RA and BN terminology

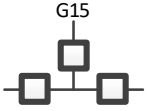
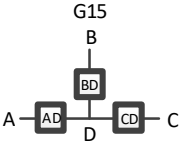
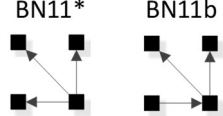
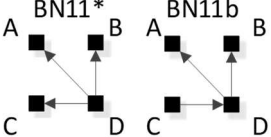
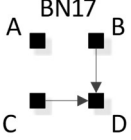
| | Our Terminology | Literature Terminology | Lattice Name, RA-like Notation | Visuals |
|----|------------------------------------|---|---|---|
| RA | General RA graph | G-structures (Klir, 1985) | G15 (Figure 7) |  |
| | Specific RA graph | Specific RA graph (Zwick, 2004) | G15 (Figure 10), <i>AD:BD:CD</i> |  |
| BN | General BN graph | Maximally oriented graphs, essential graphs, equivalence classes of directed acyclic graphs, partially directed graphs (Andersson, 1997; Chickering, 2002; Gillispie, 2001; Meek, 1995) | BN11* & BN11b (Figure 14) |  |
| | Specific BN graph (no-V-structure) | Labeled maximally oriented graphs, essential graphs, equivalence classes of directed acyclic graphs, partially directed graphs | BN11*, BN11b (Figure 20), <i>AD:BD:CD</i> |  |
| | Specific BN graph (V-structure) | acyclic graphs, partially directed graphs | BN17 (Figure 20), <i>BCD_B:C:A</i> |  |

Table 10 summarizes the RA and BN terminology and supports the discussion of BN that follows. Entries in the table for RA general and specific graphs (the lattices of general and specific graphs from Figure 7 and Figure 10, respectively) have already been discussed above. The discussion that follows this table will explain the additional

representations of BN general graphs in the insert of Figure 14, and will derive the lattice of specific BN graphs summarized in Figure 20 presented below in the section titled “Lattice of BN General and Specific Graphs.”

Additional Representations of BN General Graphs

There are 20 general graphs in the BN lattice. However, eight of these, marked with asterisks in Figure 14, namely BN2*, BN4*, BN5*, BN9*, BN11*, BN14*, BN15*, and BN16*, represent Markov equivalence classes that include additional unique edge topologies that have identical probability distributions when applied to data. These additional topologies, shown in the insert at the bottom right of Figure 14, cannot be made equivalent to the representative graphs (those with asterisks) by any 1:1 mapping of unlabeled variables. This property, described by Heckerman (1994), who showed that BNs with differing edge topologies can have the same independence structure and thus the same probability distribution, is unique to BN and is not found in RA, where there is a single unique representation of each RA general graph. All general graphs in Figure 14 without an asterisk have no Markov equivalent representations.

Two Bayesian Networks are Markov equivalent if and only if they have the same skeleton and the same V-structure (Verma, 1990), resulting in the same underlying independence structure. The skeleton of a graph is its undirected representation. As already defined, a V-structure occurs when two or more directed edges that are not themselves directly connected by an edge converge on a single node. Figure 15 shows

an example of Markov non-equivalent (Example 1) and equivalent (Example 2) BN general graphs.

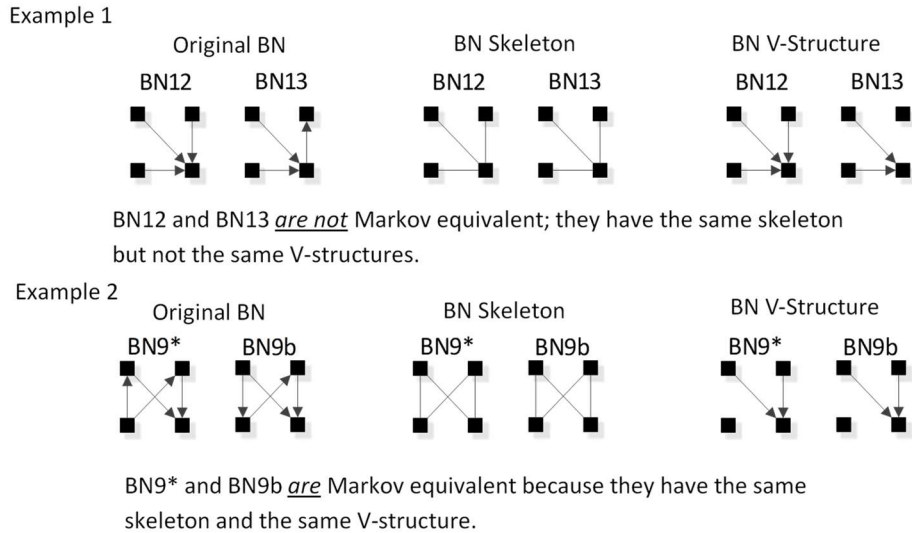


Figure 15 Examples of Markov equivalence tests

BNs that are Markov equivalent define an equivalence class; this is illustrated by BN2* in Figure 16 for which two other general graphs (BN2b and BN2c) included in the insert at the bottom of Figure 14 are in the same equivalence class. All three general graphs are Markov equivalent because they have the same skeleton and V-structures, and thus the same independence structure, but they have semantically different edge orientations. BN2* was chosen arbitrarily to represent this equivalence class and its unique independence structure. BN2b and BN2c have the same independence structure, and for corresponding variable labels, have identical probability distributions.

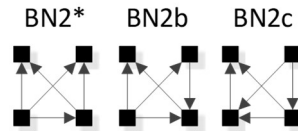


Figure 16 BN2*, BN2b, BN2c

A BN general graph is represented in the literature by an unlabeled PDAG (Chickering, 2002), also known as a Maximally Oriented Graphs (Meek, 1995), Essential Graph (Andersson, 1997) and equivalence classes of directed acyclic graphs (Gillispie, 2001). In a PDAG, edges can be directed, undirected or a mix of directed and undirected. A PDAG includes edge direction when a V-structure is present and removes edge direction when no V-structure is present. If there are no V-structures in a given BN, all edges are undirected in its PDAG representation. Figure 17 shows the PDAG representation of the graphs shown in the insert at the bottom of Figure 14. (PDAG2 encompasses BN2b and BN2c, etc.) Undirected edges can have either direction as long as a cycle is not created and also a V-structure is not created that is represented by another BN general graph. For example PDAG16, labeling variables A, B, C, D in order of left to right, top to bottom could be oriented as $B \leftarrow D \rightarrow C$ (BN16*) or $B \rightarrow D \rightarrow C$ (BN16b) (or its mirror image) but could not be oriented as $B \rightarrow D \leftarrow C$, because that creates a V-structure resulting in a different independence structure represented separately by BN17.

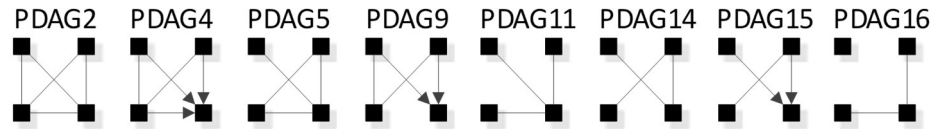


Figure 17 PDAGs for graphs in Figure 14 insert

Although representation of an entire Markov equivalence class in a single PDAG is useful, the PDAG does not visibly display the fact that semantically different edge topologies inhere in many BN general graphs (in 8 of 20 general graphs in the four-variable lattice). Use of Figure 14 to display the BN general graph lattice opts instead to show representatives of these classes and also their alternative topologies in the insert at the bottom of the figure.

BN Specific Graph Notation

A BN specific graph is simply a labeled BN general graph. As summarized in Table 10, we use the terminology of “specific graph” for what in the BN literature is called a labeled maximally oriented graph or essential graph or equivalence class of directed acyclic graphs or partially directed graph; these four different terms all refer to the same thing. All specific graphs for a given BN general graph class can be generated by permuting all possible variable labels. Given data, two BN specific graphs with different labels from the same BN general graph class will produce different probability distributions.

The notation that we use for BN specific graphs is derived from the RA notation described previously. As in RA, the colon represents marginal or conditional independence among variables and relations. For example, Figure 18, shows a labeled

version of RA general graph G15 and BN general graph BN11* which can also equivalently be represented by BN11b, both of which have the same independencies $(A \perp B, C \mid D)$, $(B \perp C \mid D)$, the same conditional probability distribution $p(A|D)p(B|D)p(C|D)p(D)$ and thus the same notation AD:BD:CD.

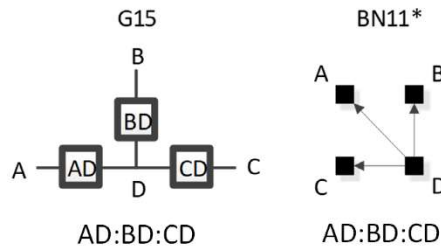


Figure 18 RA and BN notation example, without subscripts

RA notation must be modified to accommodate the V-structures that are unique to BNs and not found in RA; this is done by adding subscripts that specify the independence relations encoded by the V-structures. (For a BN graph without a V-structure, BN notation is identical to the RA notation.) For example, BN17 in Figure 19 (a) has the notation $BCD_{B:C}:A$, where the colon between $BCD_{B:C}$ and A states the independency $(A \perp B, C, D)$, namely that A is marginally independent of B , C , and D . The subscript $_{B:C}$ states marginal independence between B and C within the triadic, dependent, BCD relation. Figure 19 (b) shows the more complex BN4, which has a V-structure in which A , B , and C have arrows going to D ; this means that it has a tetradic dependency between A , B , C and D , which will be reflected in a $p(D|ABC)$ in the probability expression for this graph. The graph also has the single independency $(A \perp B \mid C)$. The notation for this graph is thus $ABCD_{AC:BC}$, which preserves the dependency

between A, B, C, and D, and also encodes the conditional independence between A and B given C. (In RA, this conditional independence is expressed by saying that $T(AC:BC) = T_c(A:B) = 0$, where T is information-theoretic transmission.)

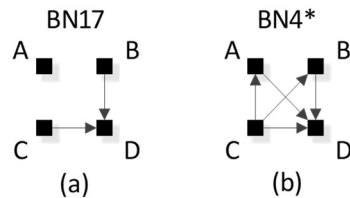


Figure 19 BN notation examples with subscripts. (a) $BCD_{B:C:A}$; (b) $ABCD_{AC:BC}$

BN Dependencies and Probability Distributions

As has been repeatedly stated in the above discussion, the marginal or conditional independence between variables and relations is what uniquely specifies an RA or BN model. “It is known that the statistical meaning of any causal model can be described economically by its stratified protocol, which is a list of independence statements that completely characterize the model” (Pearl, 1987, 1988; Verma, 1990). The method to determine BN independencies is known as D-separation, and is described in Appendix A.2. To determine the list of independence statements that completely describe any BN, D-separation is applied to *all possible* independence statements for a given BN. Those satisfying independence among variables are retained and represent the set of independencies that fully describe the structure of relations within a given BN. For four variables, Table 11 provides all possible independence statements. For a given BN, with node labels and directed edges, all independence

statements from this table need to be tested. Independence statements that are satisfied are kept, and represent the set of independencies that fully describe that BN.

Table 11 Four-variable independence statements⁸

| General Expression | Marginal Independence | | | Conditional Independence | |
|---------------------|-----------------------|--------------------------|---------------------------------|---------------------------|----------------------------------|
| | $(. \perp \dots)$ | $(. \perp \dots, \dots)$ | $(. \perp \dots, \dots, \dots)$ | $(. \perp \dots \dots)$ | $(. \perp \dots \dots, \dots)$ |
| Specific Expression | | | | | |
| 1 | $(A \perp B)$ | $(A \perp B, C)$ | $(A \perp B, C, D)$ | $(A \perp B C)$ | $(A \perp B C, D)$ |
| 2 | $(A \perp C)$ | $(A \perp B, D)$ | $(B \perp A, C, D)$ | $(A \perp B D)$ | $(A \perp C B, D)$ |
| 3 | $(A \perp D)$ | $(A \perp C, D)$ | $(C \perp A, B, D)$ | $(A \perp C D)$ | $(A \perp D B, C)$ |
| 4 | $(B \perp C)$ | $(B \perp A, C)$ | $(D \perp A, B, C)$ | $(B \perp A C)$ | $(B \perp A C, D)$ |
| 5 | $(B \perp D)$ | $(B \perp A, D)$ | | $(B \perp A D)$ | $(B \perp C A, D)$ |
| 6 | $(C \perp D)$ | $(B \perp C, D)$ | | $(B \perp C D)$ | $(B \perp D A, C)$ |
| 7 | | $(C \perp A, B)$ | | $(C \perp A B)$ | $(C \perp A B, D)$ |
| 8 | | $(C \perp A, D)$ | | $(C \perp A D)$ | $(C \perp B A, D)$ |
| 9 | | $(C \perp B, D)$ | | $(C \perp B D)$ | $(C \perp D A, B)$ |
| 10 | | $(D \perp A, B)$ | | $(D \perp A B)$ | $(D \perp A B, C)$ |
| 11 | | $(D \perp A, C)$ | | $(D \perp A C)$ | $(D \perp B A, C)$ |
| 12 | | $(D \perp B, C)$ | | $(D \perp B C)$ | $(D \perp C A, B)$ |

D-separation can also be used to test the Markov equivalence of any labeled

BNs. If two BNs have the same independencies as revealed by D-separation tests, they

⁸ The table includes typographically-different independence statements without regard to possible equivalences, thus some independencies are redundant.

are in the same Markov equivalence class and thus the same BN general graph. The prior section, however, provided a simpler way, illustrated above in Figure 15, to test for Markov equivalence of two BNs with different edge topologies.

Lattice of BN General and Specific Graphs

The BN literature on lattices predominately focuses on search algorithms to find the best BN given a scoring metric. Implicit in these search algorithms is a lattice of candidate graphs being explored in search of the best model. Chickering (2004) and others have shown the search problem to be NP-hard, with four variables there are 543 possible BNs, with 10 variables there are $O(10^{18})$ (Murphy, 1998). Because of this, research in this area has focused less on characterizing exhaustively the lattice of BN graphs, and more on advancing search heuristics to efficiently traverse the lattice to identify the best BN given a scoring metric (Bouckaert, 1994; Buntine, 1991a, 1991b; Chickering, 1995; Cooper, 1992; Friedman, 1996, 2003; Koivisto, 2004; Larranaga, 1996; Malone; and others).

Heckerman (1994) first showed that BNs with differing edge topologies can have the same independence structure and the same probability distribution, herein described as BN specific graphs. In contrast to heuristics that search all BNs, search heuristics for BN specific graphs have proven to be more efficient because they reduce the dimensionality of search space (Andersson, 1997; Chen, 2016; Chickering, 1995, 2002; Gillispie, 2001; Studený, 2010; Tian, 2010; Zhang, 2004; and others). For four variables, this approach reduces the search space from 543 BNs to 185 BN specific

graphs (Andersson, 1997). These 185 BN specific graphs can be summarized by 20 BN general graphs all with unique independence structures when variable labels are removed.

Building from the RA work of Klir (1986) and Zwick (2001), and the BN work of Pearl (1985, 1987, 1988, 2000), Verma (1990), Heckerman (1994), Chickering (1995a, 1995b, 1997, 2002, 2004), Andersson (1997), Rubin (1978), and others, the following procedure was used to generate the four variable BN general and specific graph lattice of Figure 20 in a way that can be integrated with the RA general graph lattice. While this procedure is applied in this paper to four variables, it could in principle be used for any number of variables, although of course as the number of variables increases the effort required increases exponentially.

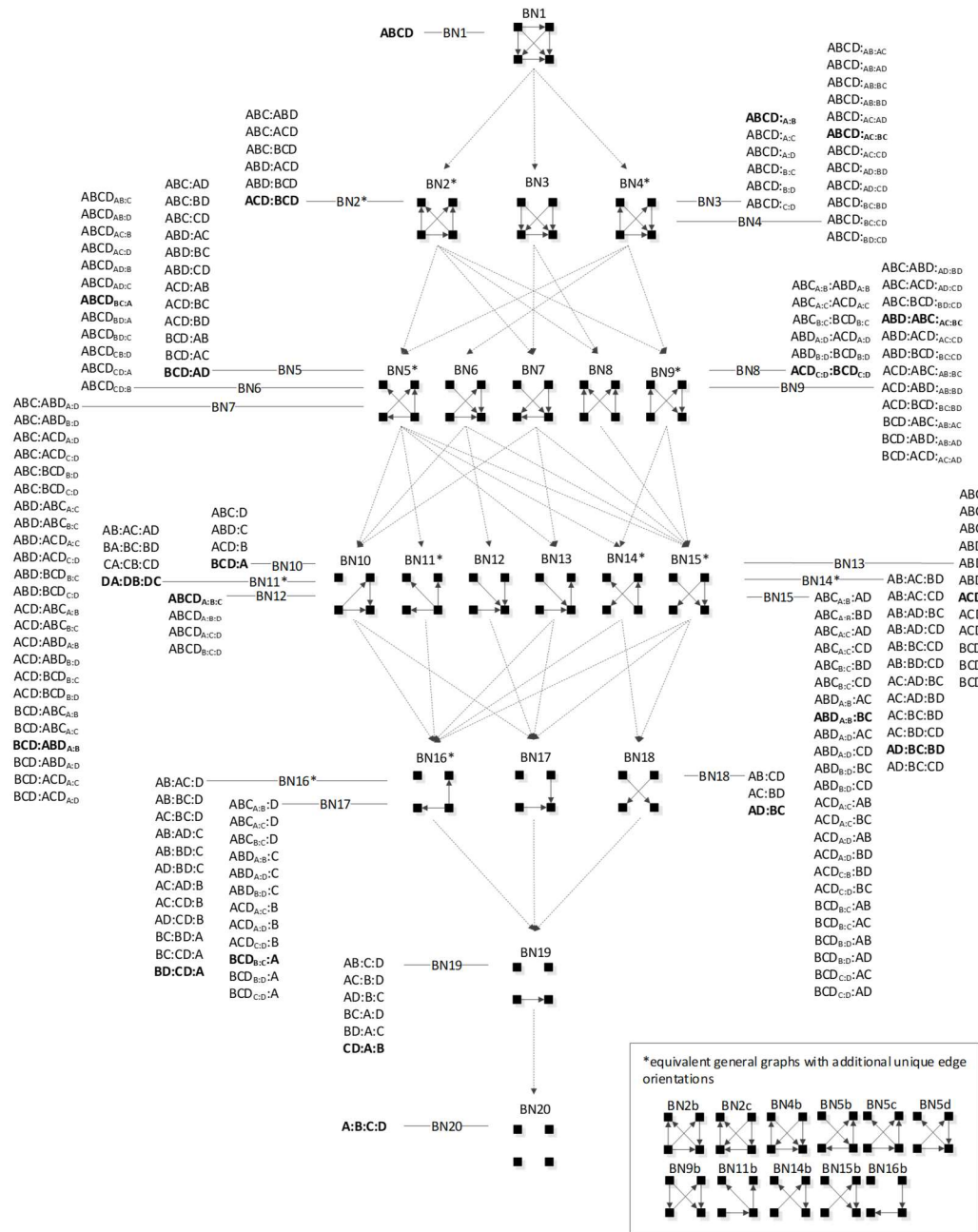


Figure 20 Lattice of general and specific BN neutral system graphs

BN Neutral System General and Specific Graph Procedure

The procedure to generate the BN neutral system general and specific graph lattice

for any number of variables is as follows:

1. Assign labels arbitrarily to the n solid squares representing variables.
2. Generate all graphs for these n variables by permuting all possible edge connections and edge orientations. Eliminate graphs with cycles. The result is the set of all labeled directed acyclic graphs for n variables.
3. For each directed acyclic graph, determine its independence structure using the D-separation procedure (MIT, 2015) detailed in Appendix A.2. This identifies which of the independence statements in Table 11 apply to the graph.
4. Collect together all graphs with the same unlabeled independencies. The set of these DAGs comprise a general graph equivalence class.
5. For each general graph equivalence class, collect together all graphs with the same labeled independencies into specific graph equivalence classes. List the RA notation for each of these specific graphs.
6. Select one specific graph equivalence class to represent the general graph, and from this specific graph equivalence class, select a single edge topology to represent the general graph. List any additional equivalent general graphs with unique edge topologies separately, as was done in the insert in Figure 14 and Figure 20.
7. Organize general graphs into levels based upon the number of edges in each general graph and link hierarchically nested general graphs in the lattice to reflect parent-child general graphs.

Figure 20 shows the result of following this procedure for four variables. This BN general and specific graph lattice can be directly compared with the RA general and specific graph lattice. The RA lattice can also be extended to include the BN lattice. The comparison and extension will be discussed in the section titled “Joint RA-BN Neutral System Lattice”.

Table 12 lists specific graph representatives for each of the general graphs in Figure 20. These specific graphs, highlighted in bold in Figure 20, assume that nodes are labeled in the order A, B, C, D from left to right, top to bottom, which is the labeling convention throughout this paper. The notation for a BN specific graph without a V-structure is identical to the RA notation. As in RA, the colon represents marginal or conditional independence among variables. For a BN graph with a V-structure, the notation adds subscripts to represent the independence relations encoded by the V-structure, which are unique to BNs and not found in RA. (See the section titled “BN specific graph notation” for more details on this notation.) Thus, graphs in Table 12 without subscripts are equivalent to an RA graph and graphs with subscripts are unique to BN. Equivalence and non-equivalence between RA and BN graphs will be discussed in the section titled “Joint RA-BN Neutral System Lattice”.

Table 12 shows for each BN general graph from Figure 20 a specific graph with its RA notation, probability distribution, and minimal list of independencies resulting from the D-separation procedure. The probability distribution is obtained as follows: (1) For each labeled node of a BN specific graph, list each node’s individual probability expression as

the probability of the node given its parents, i.e., $p(\text{node} \mid \text{parents})$; if there are no parents, simply the $p(\text{node})$. (2) Join the list of probability expressions. For example, for BN2* in Figure 21, the individual probability expressions are $p(A \mid C, D)$ for A, $p(B \mid C, D)$ for B, $p(C)$ for C, and $p(D \mid C)$ for D. Joining these gives $p(A \mid C, D)p(B \mid C, D)p(C)p(D \mid C)$. (The table omits the commas for variables that are given in conditional probability terms.)

Table 12 Probability distribution and independencies of BN specific graph examples

| BN general graph | Specific Graph Example | | |
|------------------|--|-----------------------------|--|
| | RA notation | Probability Distribution | Independencies |
| BN1 | ABCD | $p(B A)p(A)p(C AB)p(D ABC)$ | none |
| BN2 | ACD:BCD | $p(A CD)p(C)p(B CD)p(D C)$ | $(A \perp B C, D)$ |
| BN3 | ABCD _{A:B} | $p(C AB)p(A)p(B)p(D ABC)$ | $(A \perp B)$ |
| BN4 | ABCD _{AC:BC} | $p(A C)p(C)p(B C)p(D ABC)$ | $(A \perp B C)$ |
| BN5 | BCD:AD | $p(A D)p(D)p(B CD)p(C D)$ | $(A \perp B, C D)$ |
| BN6 | ABCD _{BC:A} | $p(B C)p(C)p(D ABC)p(A)$ | $(A \perp B, C)$ |
| BN7 | BCD:ABD _{A:B} | $p(C BD)p(B)p(D AB)p(A)$ | $(A \perp B), (A \perp C B, D)$ |
| BN8 | ACD _{C:D} :BCD _{C:D} | $p(A CD)p(C)p(D)p(B CD)$ | $(C \perp D), (A \perp B C, D)$ |
| BN9 | ABD:ABC _{AC:BC} | $p(A C)p(C)p(B C)p(D AB)$ | $(A \perp B C), (C \perp D A, B)$ |
| BN10 | BCD:A | $p(B C)p(C)p(D BC)$ | $(A \perp B, C, D)$ |
| BN11 | AD:BD:CD | $p(A D)p(D)p(B D)p(C D)$ | $(A \perp B, C D), (B \perp C D)$ |
| BN12 | ABCD _{A:B:C} | $p(D ABC)p(A)p(B)p(C)$ | $(A \perp B, C), (B \perp C)$ |
| BN13 | ACD _{A:C} :BD | $p(B D)p(D AC)p(A)p(C)$ | $(A \perp C), (B \perp A, C D)$ |
| BN14 | AD:BC:BD | $p(A D)p(D)p(B D)p(C B)$ | $(A \perp B D), (C \perp A, D B)$ |
| BN15 | ABD _{A:B} :BC | $p(C B)p(B)p(D AB)p(A)$ | $(A \perp B, C), (C \perp D A, B)$ |
| BN16 | BD:CD:A | $p(B D)p(D)p(C D)p(A)$ | $(B \perp C D), (A \perp B, C, D)$ |
| BN17 | BCD _{B:C} :A | $p(D BC)p(B)p(C)p(A)$ | $(B \perp C), (A \perp B, C, D)$ |
| BN18 | AD:BC | $p(C B)p(B)p(D A)p(A)$ | $(A, D \perp B, C)$ |
| BN19 | CD:A:B | $p(D C)p(C)p(A)p(B)$ | $(B \perp C, D), (A \perp B, C, D)$ |
| BN20 | A:B:C:D | $p(A)p(B)p(C)p(D)$ | $(A \perp B, C, D), (B \perp C, D), (C \perp D)$ |

$$p(A|C,D)p(B|C,D)p(C)p(D|C)$$

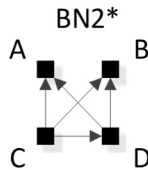


Figure 21 Probability distribution for BN2* example

The equivalence or non-equivalence of RA and BN graphs is discussed in detail in the section titled “Joint RA-BN Neutral System Lattice”, below, but Table 12 provides an advanced look at this issue. Any BN general graph with a specific graph example whose RA notation does not include subscripts is equivalent to some general RA graph; there are 10 of these BN general graphs. Any BN general graph with a specific graph example whose notation includes subscripts is not equivalent to any general RA graph; there are also 10 of these BN general graphs, which all have V-structures.

BN Directed Systems

The BN discussion so far has focused on BN neutral systems in which an IV-DV distinction is not made. This section narrows the focus to BN predictive graphs, analogous to RA directed systems, where the aim is to predict a single DV given the IVs. As in RA, we define Z as the dependent variable in the BN directed system lattice, replacing variable D in the neutral system lattice. We designate as the DV in a given BN any node with the exception of a parent node within a V-structure. That is, we do not consider here the possibility that a parent node within a V-structure could be designated as a DV; this will be discussed further in the section titled “Comparing RA and BN directed system graphs”. As is the case for RA, many graphs in the neutral system lattice are redundant when the aim is only to predict the DV. The BN directed system lattice of Figure 22, where only graphs with unique predictions of Z are highlighted, is thus a subset of the BN neutral system lattice of Figure 20. For each general graph in Figure 22 with a unique prediction, associated specific graphs are listed.

Specific graphs that are bolded correspond to the displayed BN edge orientation and edge connections assuming labeling of nodes from top left, top right, bottom left, bottom right as A, B, C, Z respectively. These bolded specific graphs also correspond to the examples below in Table 13. Graphs not highlighted in Figure 22 are equivalent in their predictions to highlighted graphs. (Asterisks in this figure have the same meaning they have in BN Figure 14 and Figure 20.) For two graphs with identical predictions, the graph with the least degrees of freedom was selected. There are eight general graphs and 18 specific graphs in the BN directed system lattice; this is a significant compression of the BN neutral system lattice that includes 20 general graphs and 185 specific graphs.

Table 13 BN directed system graphs

| BN general graph | Predictively equivalent simpler graph | Specific graph example RA notation | Specific graph example Probability distribution |
|-------------------------|--|--|--|
| BN1 | BN12 | ABCZ | $p(Z ABC)p(C AB)p(B A)p(A)$ |
| BN2 | BN7 BN17 | ABZ:BCZ ABC:BCZ | $p(C BZ)p(Z AB)p(A B)p(B)$ $p(Z BC)p(B CA)p(A C)p(C)$ |
| BN3 | BN12 | ABCZ _{A:B} | $p(Z ABC)p(C AB)p(A)p(B)$ |
| BN4 | BN12 | ABCZ _{AC:BC} | $p(Z ABC)p(A C)p(B C)p(C)$ |
| BN5 | BN13 BN19 | ACZ:BZ ABC:CZ | $p(Z AC)p(B Z)p(C A)p(A)$ $p(Z C)p(B CA)p(C A)p(A)$ |
| BN6 | BN12 | ABCZ _{BC:A} | $p(Z ABC)p(B C)p(C)p(A)$ |
| BN7 | BN17 | BCZ:ABZ _{A:B} BCZ:ABC _{A:C} | $p(C BZ)p(Z AB)p(B)p(A)$ $p(Z BC)p(B CA)p(C)p(A)$ |

| | | | |
|------|--------------|----------------------------------|--|
| BN8 | BN17 | $ABC_{B:c}:BCZ_{B:c}$ | $p(Z BC)p(A BC)p(B)p(C)$ |
| BN9 | BN17 | $BCZ:ABC_{AB:AC}$ | $p(Z BC)p(B A)p(C A)p(A)$ |
| BN10 | BN17 | $BCZ:A$ | $p(Z BC)p(B C)p(C)p(A)$ |
| BN11 | BN19 | $AZ:BZ:CZ$ $AC:BC:CZ$ | $p(A Z)p(B Z)p(C Z)p(Z)$ $p(Z C)p(B C)p(C A)p(A)$ |
| BN12 | | $ABCZ_{A:B:c}$ | $p(Z ABC)p(A)p(B)p(C)$ |
| BN13 | BN19 | $ACZ_{A:c}:BZ$ $ABC_{A:B}:CZ$ | $p(Z AC)p(B Z)p(A)p(C)$ $p(Z C)p(C AB)p(A)p(B)$ |
| BN14 | BN16 BN19 | $AB:BZ:CZ$ $AB:BC:CZ$ | $p(Z B)p(C Z)p(B A)p(A)$ $p(Z C)p(B A)p(A)p(C B)$ |
| BN15 | BN17 BN19 | $BCZ_{B:c}:AB$ $ABC_{A:c}:CZ$ | $p(Z BC)p(A B)p(B)p(C)$ $p(Z C)p(B CA)p(A)p(C)$ |
| BN16 | BN19 | $BZ:CZ:A$ $BC:CZ:A$ | $p(B Z)p(C Z)p(Z)p(A)$ $p(Z C)p(C B)p(B)$ |
| BN17 | | $BCZ_{B:c}:A$ | $p(Z BC)p(B)p(C)p(A)$ |
| BN18 | BN19 | $AB:CZ$ | $p(Z C)p(B A)p(A)p(C)$ |
| BN19 | | $CZ:A:B$ | $p(Z C)p(C)p(A)p(B)$ |
| BN20 | | $A:B:C:Z$ | $p(Z)p(A)p(B)p(C)$ |

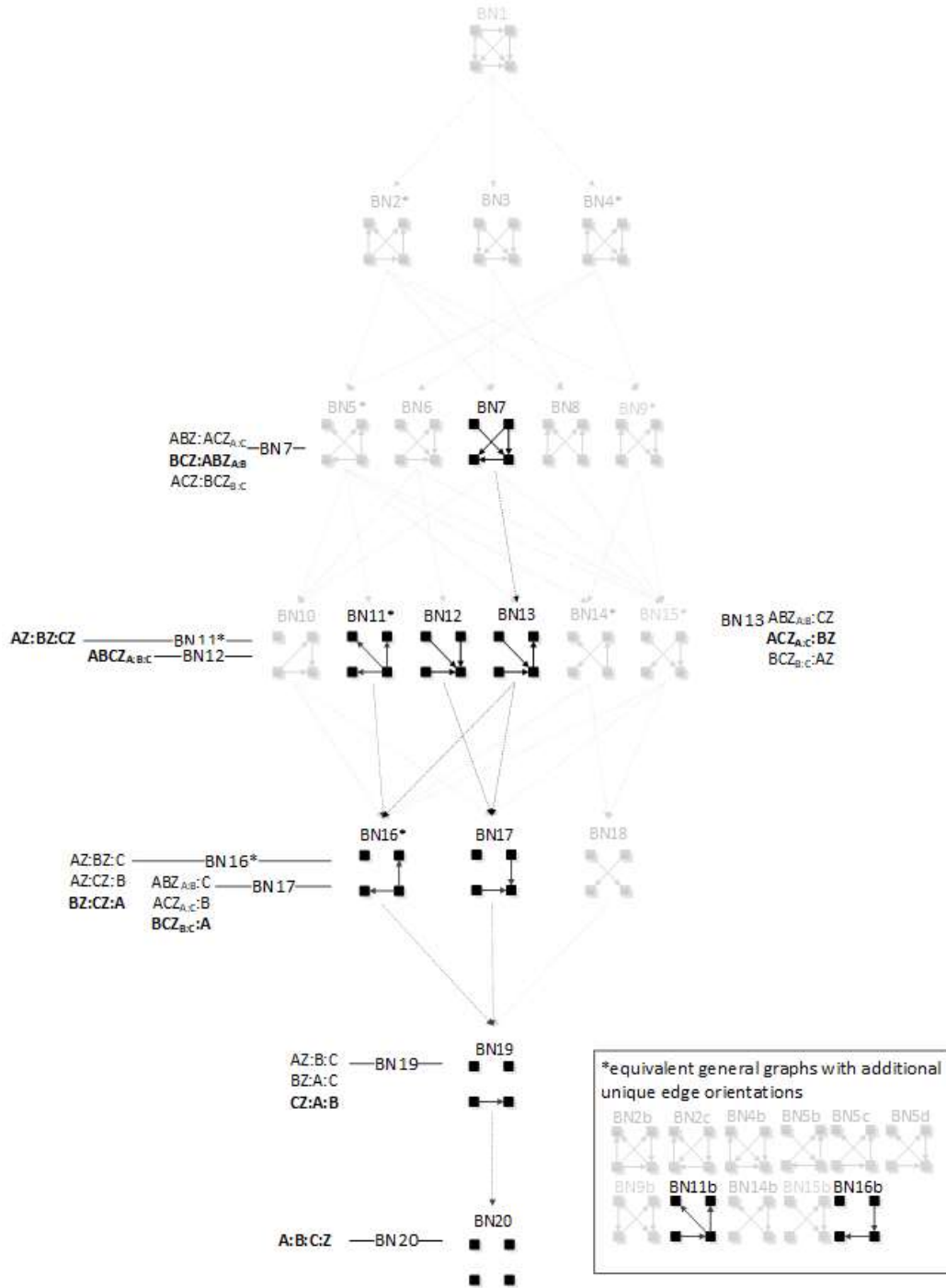


Figure 22 BN directed system lattice

Table 13 lists all BN directed system general graphs. When BN graphs are greyed in column 1 it means the graph is equivalent in terms of prediction to a simpler (fewer

degrees of freedom) general graph. Column 2 identifies which simpler graph it is equivalent to. General graphs with a blank row in column 2 have no simpler equivalently predicting graph, and are included in the directed system lattice of Figure 22. Column 3 provides specific graph examples of these general graphs and column 4 shows the specific graph probability distributions. Within column 4, only the expressions that are used to predict the dependent variable are highlighted in black. All other non-predictive relations are greyed. For example, BN1, BN3, BN4, and BN6 and BN12 all predict Z in the same way, i.e., $p(Z|ABC)$, thus they are all equivalent in terms of prediction. However, BN12 has the least degrees of freedom and is therefore selected to represent all five of these equivalent general graphs.

Joint RA-BN Neutral System Lattice

Joint RA-BN Neutral System Lattice Introduction

This section integrates the RA and BN neutral system general graph lattices using the four variable Rho lattice (Klir, 1985). Combining the Rho, RA and BN lattice creates a larger and more descriptive lattice than any previously identified in the literature. The lattice identifies independence structures unique to RA or to BNs, and independence structures that are equivalent across RA and BN. Equivalence is in terms of independence structure as described separately for RA in the section titled “RA Lattice” and BN in the section titled “BN Lattice”. Where two or more graphs, RA or BN, have the same general independence structure regardless of variable labels, they are equivalent. General independence structure is represented with independence statements without

labels. For example, $(. \perp . . | \dots)$, one variable is independent of another, given a third. Consider, for example, RA general graph G15 and BN general graph BN11 have the same general independence structure $(. \perp . . , \dots | \dots)$, $(. . \perp . . | \dots)$, thus they are equivalent. Two specific graphs are equivalent if they have the same independence structure given variable labels. For example, using RA general graph G15 and BN general graph BN11 again, Figure 23 shows these general graphs with variable labels added making them specific graphs. Given these labels, they have equivalent general and specific independence structure, $(. \perp . . , \dots | \dots)$, $(. . \perp . . | \dots)$ and $(A \perp B, C | D)$, $(B \perp C | D)$ respectively.

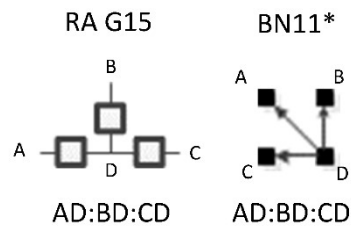


Figure 23 G15 and BN11* specific graph example

RA-BN Rho Neutral System Graphs

The Rho (ρ) lattice of Figure 24 (adapted from Klir, 1985, p. 237) is a simplification of the RA lattice of general graphs and is used here to integrate the RA neutral system lattice with the BN neutral system lattice. The Rho lattice is an even more general lattice than the RA general graph lattice and can map both RA and BN general graphs to one of its eleven structures. A solid dot represents a variable; a line connects variables in the Rho lattice if these two variables are directly connected by any box (relation) in the RA general graph lattice. Arrows from one Rho graph to another

represent hierarchy, i.e., the generation of a child graph from a parent graph. ρ_1 represents maximal connectedness, or dependence, between variables, and ρ_{11} represents independence among all variables. Graphs in-between ρ_1 and ρ_{11} represent a mix of dependence and independence among variables. Each RA or BN general or specific graph corresponds to one, and only one, of the eleven Rho graphs.

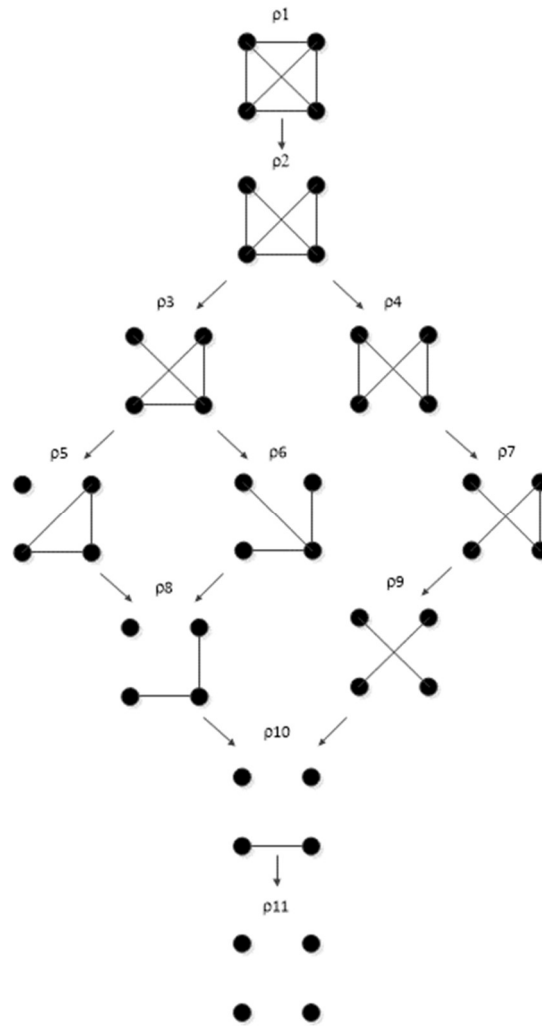


Figure 24 Lattice of four-variable Rho graphs

Rho and Equivalent RA and BN General Graphs

Out of 20 RA neutral system general graphs and 20 BN neutral system general graphs, there are 10 RA general graphs, comprising all of the graphs with no loops in the RA lattice that are equivalent to BN general graphs. Each of these RA-BN equivalent pairs corresponds to one of the 11 Rho graphs from Figure 24, with the exception of ρ_4 . ρ_4 has corresponding RA and BN general graphs, but these do not have equivalent independence structures, and are discussed in the following the section titled “Rho and non-equivalent RA and BN general graphs”.

ρ_1 reflects maximal connectedness among all four variables. For both the RA general graph G1 and the BN general graph BN1 from Figure 7 and Figure 14 respectively, there are no independencies among the variables and thus the graphs are equivalent. Both graphs have only one specific graph, ABCD. This is summarized in Figure 25.

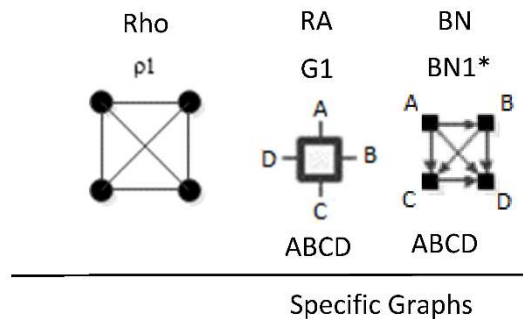


Figure 25 Rho1, G1 and BN1 specific graph

ρ_2 corresponds to RA general graph G7 and BN general graph BN2*, as shown in Figure 26. It is clear how BN2* corresponds to Rho graph ρ_2 because visually they are

represented in the same way with the exception that the Rho graph has undirected edges. There are two additional BN general graphs (BN3 and BN4*) that correspond to ρ_2 ; however they have no equivalent RA general graph, so they are discussed in the next section which concerns non-equivalent RA and BN general graphs. ρ_2 , G7, and BN2* represent two three-variable relations with conditional independence between two variables, with general independence structure $(. \perp . . | \dots, \dots)$. Assigning labels to variables makes it easier to interpret the RA association with ρ_2 . Figure 26 shows an example with variable labels (one of six possible permutations of variable labels) assigned to RA graph G7 which results in RA specific graph ACD:BCD, in which A is independent of B given C and D, $(A \perp B | C, D)$. Assigning labels to the BN graph in Figure 26 yields the same specific graph. Other label permutations yield five other equivalent RA and BN specific graphs: ABC:ABD, ABC:ACD, ABC:BCD, ABD:ACD, ABD:BCD.

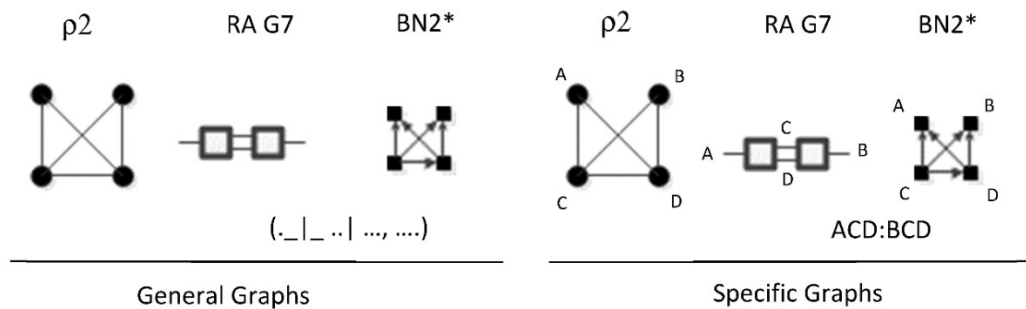


Figure 26 Rho2, G7, and BN2* example

ρ_3 represents RA graph G10 and BN graph BN5* which have the same independence structure, $(. \perp . . , \dots | \dots)$. Figure 27 shows an example of one of eight RA

G10 and BN5* specific graphs, BCD:AD, with independencies ($A \perp B, C \mid D$). The full list of eight specific RA G10 and BN5* specific graphs are: ABC:AD, ABC:BD, ABC:CD, ABD:AC, ABD:BC, ABD:DC, ACD:AB, ACD:CB, ACD:DB, BCD:BA, BCD:CA, and BCD:DA. G10 has been previously characterized as naïve BN-like.

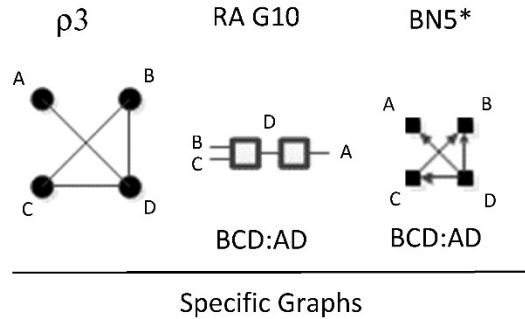


Figure 27 Rho3, G10, and BN5* specific graph example

ρ_4 is discussed later in the section on non-equivalent RA and BN general graphs.

ρ_5 represents RA graph G13 and BN graph BN10 which have the same independence structure, ($A \perp B, C, D$) in that they have no independencies in the triadic relation and the fourth variable is independent of all three variables in the triadic relation. Figure 28 shows an example of one of four RA G13 and BN10 specific graphs, BCD:A, with independencies ($A \perp B, C, D$). The full list of RA G13 and BN10 specific graphs are: ABC:D, ABD:C, ACD:B, and BCD:A.

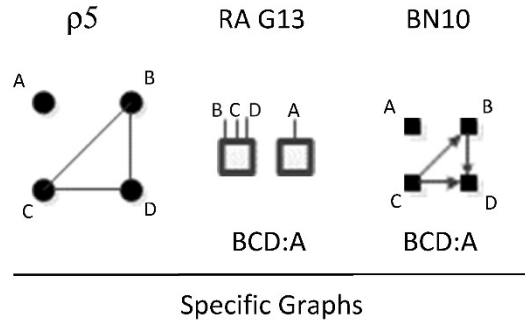


Figure 28 Rho5, G13, and BN10* specific graph example

ρ_6 represents RA general graph G15 and BN graph BN11* which have the same independence structure, $(. \perp \dots, \dots \mid \dots)$, $(\dots \perp \dots \mid \dots)$. There are three dyadic relations in these graphs with one variable present in all three dyadic relations and the other three variables present in only one of three dyadic relations.

This graph is described in the literature (Zhang, 2004) as a naïve BN, simple Bayes, or independence Bayes, because of its simple dyadic relations among variables. What is also clear is RA general graph G15 represents a naïve BN because of its equivalent independence structure. Figure 29 shows an example of one of RA G15 and BN11* specific graphs, AD:BD:CD, with independencies $(A \perp B, C \mid D)$, $(B \perp C \mid D)$, and conditional probability distribution $p(A \mid D) p(B \mid D) p(C \mid D) p(D)$. The full list of specific RA G15 and BN11* specific graphs are: AB:AC:AD, AB:BC:BD, AC:BC:CD, and AD:BD:CD.

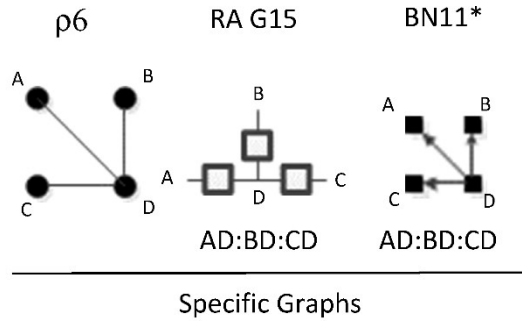


Figure 29 Rho 6, G15 and BN11* specific graph example

ρ_7 represents RA graph G16 and BN graph BN14* which have the same independence structure, $(. \perp .. | \dots)$, $(... \perp ., \dots | ..)$. Figure 30 shows an example of one of twelve RA G16 and BN14* specific graphs, AD:BC:BD, with independencies $(A \perp B | D)$, $(C \perp A, D | B)$ and conditional probability distribution $p(A|D)p(B|D)p(C|B)p(D)$. The full list of specific RA G16 and BN14* specific graphs are: AB:AC:BD, AB:AC:CD, AB:AD:BC, AB:AD:CD, AB:BC:CD, AB:BD:CD, AC:AD:BC, AC:AD:BD, AC:BC:BD, AC:BD:CD, AD:BC:BD, and AD:BC:CD.

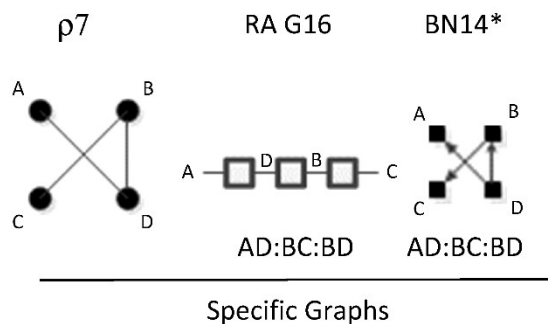


Figure 30 Rho7, G16 and BN14* specific graph example

ρ_8 represents RA general graph G17 and BN general graph BN16* which have the same independence structure, $(.. \perp ... | \dots)$, $(. \perp .., \dots, \dots)$. There are two dyadic relations in these graphs with one variable present in both dyadic relations, and the

fourth variable not present in either dyadic relation, and thus independent of the three other variables. This graph is also representative of a naïve BN. Figure 31 shows an example of one of twelve RA G17 and BN16* specific graphs, BD:CD:A, with independencies $(B \perp C \mid D), (A \perp B, C, D)$, and conditional probability distribution $p(B|D) p(C|D)p(D)$. The full list of specific RA G17 and BN16* specific graphs are: AB:AC:D, AB:BC:D, AC:BC:D, AB:AD:C, AB:BD:C, AD:BD:C, AC:AD:B, AC:CD:B, AD:CD:B, BC:BD:A, BC:CD:A, and BD:CD:A.

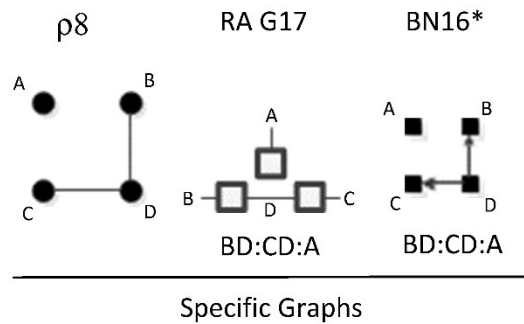


Figure 31 Rho 8, G17 and BN16* specific graph example

ρ_9 represents RA general graph G18 and BN general graph BN18 which have the same independence structure, $(. , \dots \perp \dots, \dots)$. There are two dyadic relations in these graphs with two variables included in one dyadic relation and the other two included in the other. Figure 32 shows an example of one of three RA G18 and BN18 specific graphs, AD:BC, with independencies $(A, D \perp B, C)$, and conditional probability distribution $p(C|B)p(B)p(D|A)p(A)$. The full list of specific RA G18 and BN18 specific graphs are: AB:CD, AC:BD, and AD:BC.

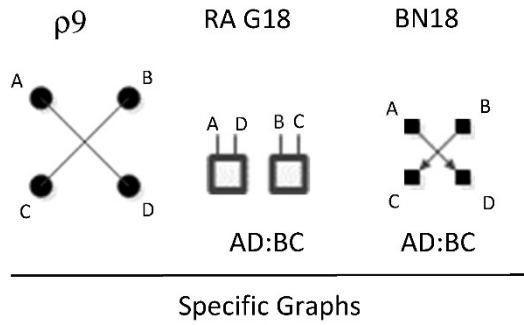


Figure 32 Rho 9, G18 and BN18* specific graph example

ρ_{10} represents RA graph G19 and BN graph BN19 have the same independence structure, $(\dots \perp \dots, \dots)$, $(\dots \perp \dots, \dots, \dots)$. There is one dyadic relation and two variables independent of all other variables. Figure 33 shows an example of one of six RA G19 and BN19 specific graphs, CD:A:B, with independencies $(B \perp C, D)$, $(A \perp B, C, D)$ and conditional probability distribution $p(D|C)p(C)p(A)p(B)$. The full list of specific RA G19 and BN19 specific graphs are: AB:C:D, AC:B:D, AD:B:C, BC:A:D, BD:A:C, and CD:A:B.

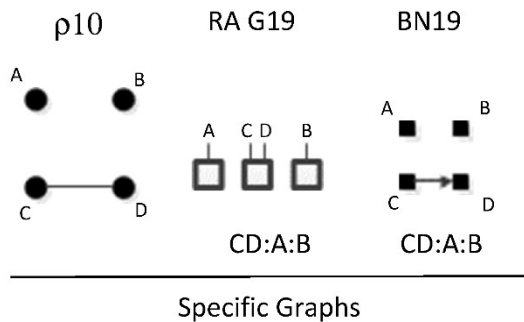


Figure 33 Rho, 10, G19 and BN19 specific graph example

ρ_{11} represents RA graph G20 and BN graph BN20 which have the same independence structure $(\dots \perp \dots, \dots)$, $(\dots \perp \dots, \dots)$, $(\dots \perp \dots)$ in which all variables are

independent of one another, $(A \perp B, C, D), (B \perp C, D), (C \perp D)$. Figure 34 shows the only specific graph for RA G20 and BN20, A:B:C:D.

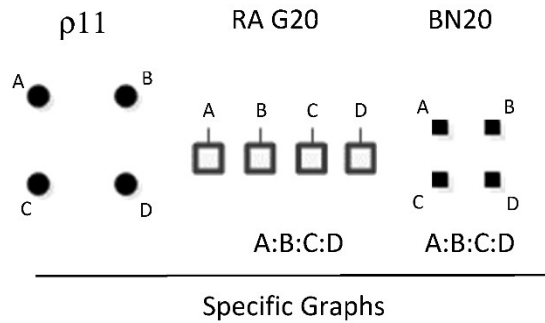


Figure 34 Rho 11, G20 and BN20 specific graph

Table 14 summarizes all equivalent RA and BN general graphs, with their associated Rho graph, an example of their specific graph notations and their independences. These specific graph examples align with the BN general graphs of Figure 14 assuming labeling of nodes A, B, C, D in the order of top left, top right, bottom left, bottom right.

Table 14 Equivalent Rho, RA and BN neutral system general graphs

| Rho | RA | BN | Specific Graph | Independencies |
|-------------|-----|-------|----------------|--|
| ρ_1 | G1 | BN1 | ABCD | no independencies |
| ρ_2 | G7 | BN2* | ACD:BCD | $(A \perp B \mid C, D)$ |
| ρ_3 | G10 | BN5* | BCD:AD | $(A \perp B, C \mid D)$ |
| ρ_5 | G13 | BN10 | BCD:A | $(A \perp B, C, D)$ |
| ρ_6 | G15 | BN11* | AD:BD:CD | $(A \perp B, C \mid D), (B \perp C \mid D)$ |
| ρ_7 | G16 | BN14* | AD:BC:BD | $(A \perp B \mid D), (C \perp A, D \mid B)$ |
| ρ_8 | G17 | BN16* | BD:CD:A | $(B \perp C \mid D), (A \perp B, C, D)$ |
| ρ_9 | G18 | BN18 | AD:BC | $(A, D \perp B, C)$ |
| ρ_{10} | G19 | BN19 | CD:A:B | $(B \perp C, D), (A \perp B, C, D)$ |
| ρ_{11} | G20 | BN20 | A:B:C:D | $(A \perp B, C, D), (B \perp C, D), (C \perp D)$ |

Rho and Non-Equivalent RA and BN General Graphs

In addition to the 10 equivalent RA and BN general graphs, there are 10 general graphs unique to the RA lattice and 10 general graphs unique to the BN lattice. All 10 non-equivalent RA general graphs in the four variable lattice have loops and require iteration to generate their probability distributions. BNs are acyclic and have analytic solutions, so there are no BN general graphs that are equivalent to the RA graphs with loops. Since RA graphs are undirected, one might think that there could be some equivalent acyclic directed BN graphs, but this is not the case, because BN graphs that are acyclic when directions are considered but cyclic if directions are ignored have V-structure interpretations, as described previously. All 10 non-equivalent BN general graphs have such V-structures, which encode independence relations unique to BNs. To illustrate: the structure $A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A$ is cyclic and not a legitimate BN structure, but the directed structure of $A \rightarrow B, B \rightarrow C, C \rightarrow D, A \rightarrow D$ (BN9b from Figure 14), which has the same undirected links, is not cyclic, and is a legitimate BN structure. However, this latter structure is not interpreted as a set of dyadic relations, which would be written in RA notation as $AB:BC:CD:AD$ and contains a loop (RA general graph G12 from Figure 7). Rather, the V-structure consisting of $C \rightarrow D$ and $A \rightarrow D$ is interpreted as a triadic relation, which contributes a $p(D|AC)$ to the probability expression, $p(A)p(B|A)p(C|B) p(D|AC)$, which does not correspond to any RA structure.

Lattice of Rho, RA, BN Neutral System General Graphs

The lattice of Rho, RA and BN equivalent and non-equivalent general graphs in Figure 35 was developed from the RA lattice in Figure 7 and the BN lattice in Figure 14. This lattice includes all 10 unique RA general graphs, 10 unique BN general graphs, and 10 RA and BN equivalent general graphs, for a total of 30 unique general graphs. The lattice is organized using the Rho lattice (Klir, 1985). All 20 RA general graphs and all 20 BN general graphs for each Rho graph are represented in the joint lattice. Within each Rho graph, where RA and BN graphs are equivalent, that is, when their independence structures are identical, the BN graph is placed under the RA equivalent graph. Where RA or BN graphs are not equivalent, representing an independence structure unique to RA or BN, they stand alone.

Arrows from one graph to another in the joint lattice represent the hierarchy of the RA lattice only. As can be seen in the section titled “BN Lattice”, the hierarchy of the BN lattice has many more links from parent to child graphs and thus is not a useful representation in the joint lattice. Additionally, Figure 40 includes the Joint RA-BN lattice of general and specific graphs. This lattice shows 53 unique RA specific graphs, 124 unique BN specific graphs, and 61 RA-BN equivalent specific graphs, for a total of 238 combined, unique, RA and BN specific graphs.

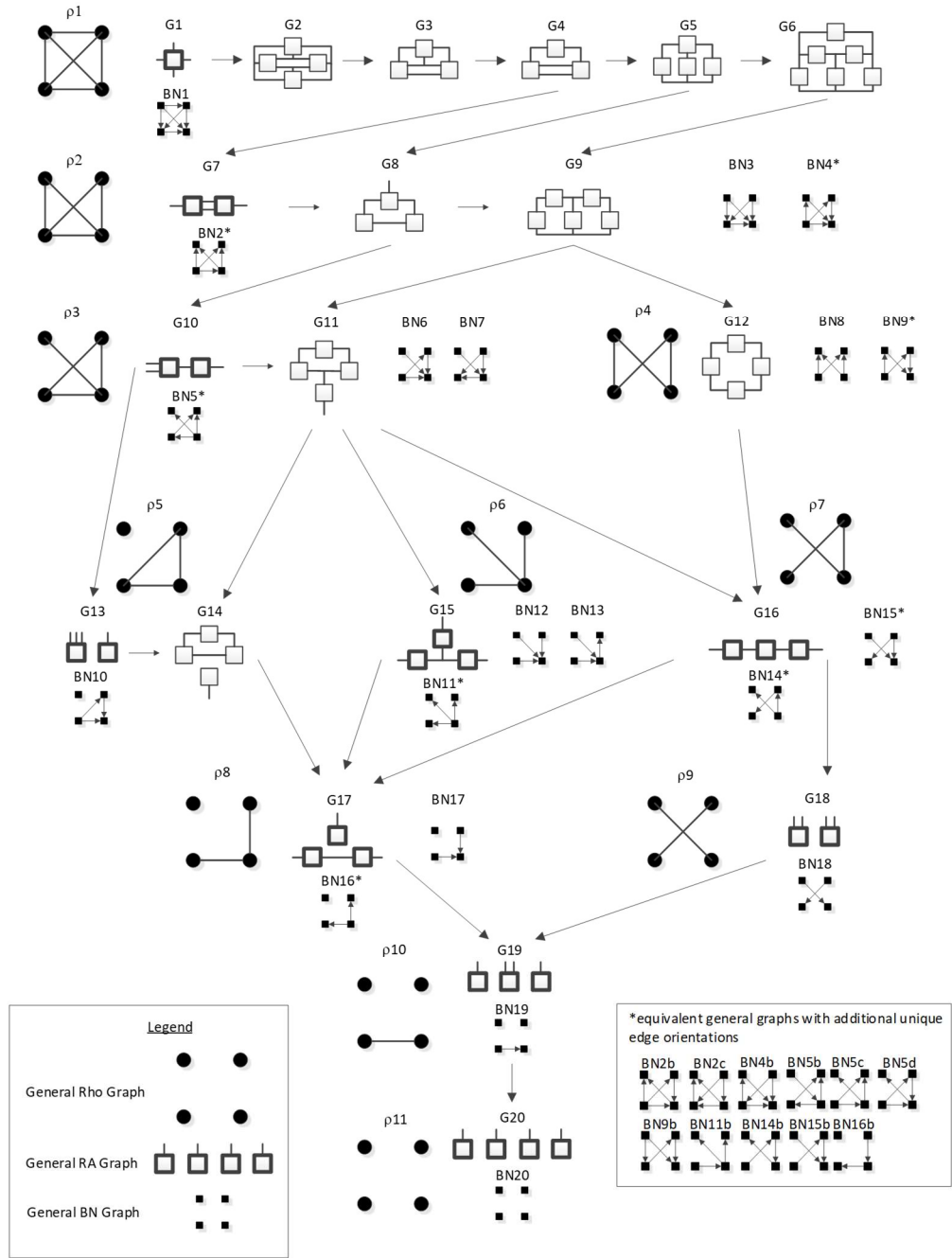


Figure 35 Lattice of 4-variable general Rho, RA and BN neutral system graphs

Joint RA-BN Lattice Algorithm

This section defines an algorithm for generating the Joint RA-BN lattice of neutral system general and specific graphs.

Procedure to Generate the RA Neutral System General and Specific Graphs from a Single Rho Graph

This is done in three steps: in Step 1, generate the most complex set of specific graphs that correspond to the Rho graph; in Step 2, generate all their less complex specific graph descendants; in Step 3, specific graphs are collected together in general graphs.

Step 1 begins with (Step 1.1) labeling the Rho graph, as shown in Figure 36. The most complex specific graph that corresponds to this labeled Rho graph is obtained (Step 1.2) by representing each clique⁹ with a single relation encompassing all the variables in the clique and then joining these relations with a ":". For example, in Figure 36, A, B, and C are in a clique, i.e., are fully linked to one another and this is also the case for B, C, and D, but A and B are not linked. The resulting specific graph is ABC:BCD, which is encompassed in RA general graph G7. Next (Step 1.3), permute all the variables in this specific graph, which generates the other five specific graphs that are encompassed within G7, as shown in RA lattice of Figure 10.

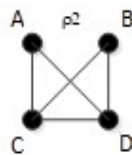


Figure 36 Example, Rho 2

⁹ "Cliques" here refer to "maximal cliques."

Step 2 then generates the simpler RA representations of G_7 that map to ρ_2 , namely the specific graphs that are encompassed within the RA general graphs G_8 and G_9 . Klir (1985, p.231) details the procedure for this step. In Step 3, specific graphs with the same independence structure are then collected together in general graph equivalence classes. Doing this for ρ_2 results in general graphs G_7 , G_8 and G_9 and their specific graphs as shown in Figure 10.

Procedure to Generate the BN Neutral System General and Specific Graphs from a Single Rho Graph

In contrast to RA graphs, BNs are just Rho graphs with directions added to edges, as shown in Figure 37. To generate all BN specific graphs for a given Rho graph, simply permute all possible edge directions and variable combinations, and follow the BN neutral system general and specific graph procedure outlined above in the section titled “BN Neutral System General and Specific Graph Procedure”. Essentially, the process entails discarding redundant specific graphs and graphs with cycles from all these permutations, and collecting together BN specific graphs with unique independence structures into a general graph.

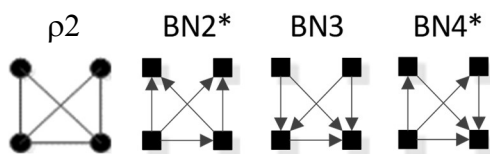


Figure 37 Rho 2 example, with associated BNs general graphs

Generating the Joint RA-BN General and Specific Graph Lattice

The following provides a general algorithm to generate the joint RA-BN lattice of neutral system general and specific graphs for any number of variables from some specific starting graph, either downwards or upwards.

1. Identify a starting Rho graph
2. Generate all possible RA and BN specific graphs for the given Rho graph.
 - a. For RA, follow the procedure detailed in the prior section titled “Procedure to generate the RA neutral system general and specific graphs from a given Rho graph”
 - b. For BN, follow the procedure detailed in the prior section titled “Procedure to generate the BN neutral system general and specific graphs from a given Rho graph”
 - c. Organize all RA and BN general graph equivalence classes into three categories: RA graphs with loops, BN graphs with V-Structures, and equivalent RA-BN graphs containing no loops or V-structures.
3. If searching the lattice upward, add an edge to the prior Rho graph. If searching the lattice downward, delete an edge from the prior Rho graph.
4. Repeat steps 2 and 3 until the top or bottom of the lattice is reached.

Consider for example the results of the RA and BN procedures for Rho 2. Organizing these results via step 2c gives the following six general structures: G8 and G9 for RA graphs with loops, BN3 and BN4* for BN graphs with V-structures, and G7 and BN2* for

equivalent RA-BN graphs. Specific structures can be simply obtained from these general structures by listing all permutations of variable labels. Following these procedures for any number of variables will result in the exhaustive, non-redundant, lattice of joint RA-BN neutral system general and specific graphs.

Comparing RA and BN Directed System Graphs

Figure 38 shows side-by-side for comparison the RA augmented directed system lattice from Figure 13 and the BN directed system lattice from Figure 22. To the left or right of each BN directed system general graph is the equivalent RA directed system general graph. For example, BN7 is equivalent to RA general graph G7. Equivalence in this context is in terms statistical equivalence of prediction results given data. Two directed system general graphs are equivalent if they predict the DV (Z) in the same way. Each of the BN directed system general graphs in the lattice is equivalent to an RA general graph in the augmented RA directed system general graph lattice. In addition, the RA directed system lattice includes additional predictive graphs, those with loops that are not found in the BN lattice. Thus, restricting BN directed systems to those where the DV is not a parent in a V-structure, the RA augmented directed system lattice fully encompasses the BN directed system lattice and offers additional predictive graphs.

Lattice of 4 Variable Augmented Directed System RA General and Specific Graphs
12 General Graphs, 30 Specific Graphs

Lattice of 4 Variable BN Directed System General and Specific Graphs
8 General Structures, 18 Specific Structures

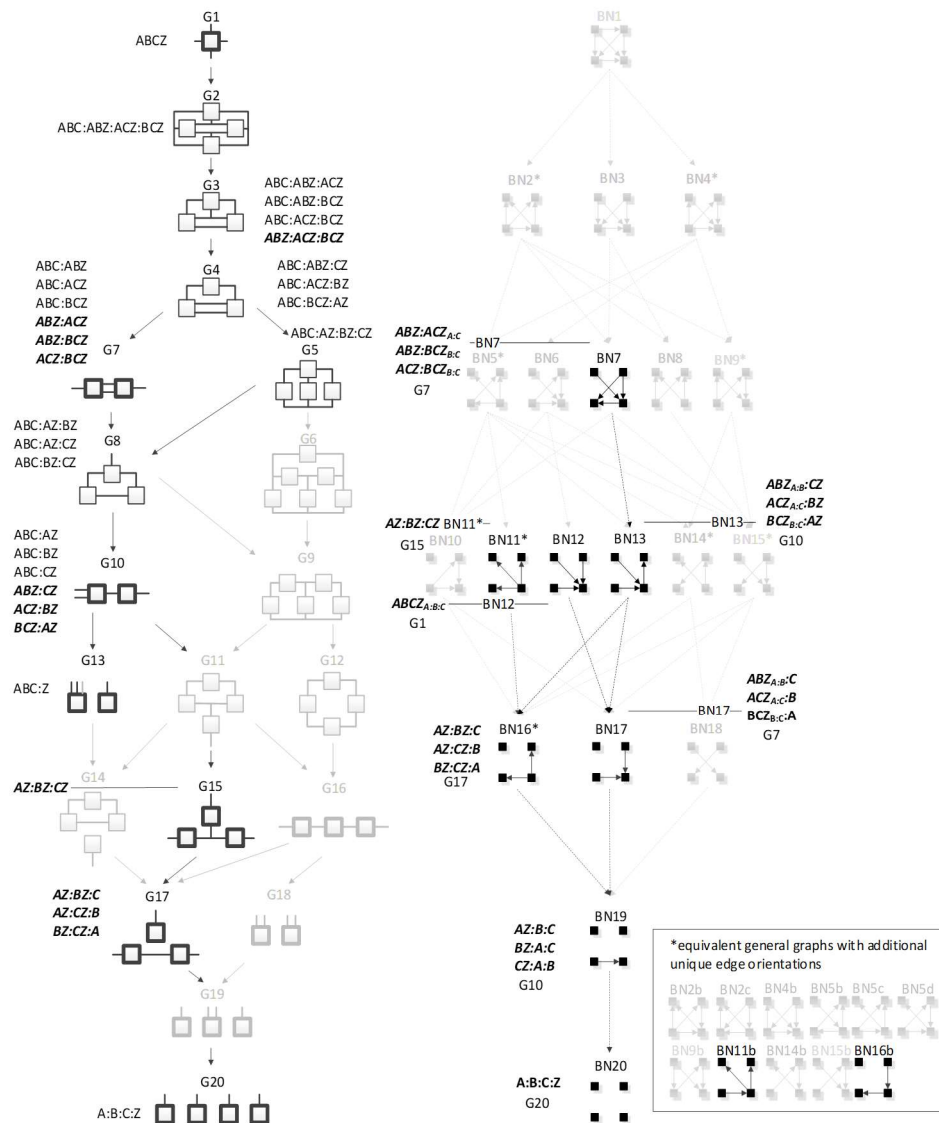


Figure 38 RA and BN directed system lattice comparison

Table 15 shows all BN directed system general graphs and their RA equivalents as well as specific graph examples with their associated probability distributions. In these probability distributions, only the terms used to predict the DV (Z) are highlighted

in black; non-predictive terms are greyed. All equivalences necessarily involve loopless RA models; half of these involve RA graphs in the standard directed system lattice, where every model has an IV component, and the other half involve graphs in the augmentation of this lattice. Prior to development of the BN directed system lattice in this paper, the RA directed system lattice did not include naïve Bayes equivalent graphs, e.g., G15 and G17, and the naïve Bayes-like graph, G10. The development of the BN directed system lattice in this paper in part inspired the augmentation of the standard RA directed system lattice to include naïve Bayes type graphs.

Table 15 BN directed system graphs and RA equivalent example

| BN general graph | BN Specific graph example RA notation | BN Specific graph example Probability distribution | Equivalent RA graph | Equivalent RA graph notation | Equivalent RA graph Probability distribution |
|-------------------------|--|---|----------------------------|-------------------------------------|---|
| BN7 | BCZ:ABZ _{A,B} | $p(C BZ)p(Z AB)p(B)p(A)$ | G7 (augmentation) | BCZ:ABZ | $p(C BZ)p(Z AB)p(B A)p(A)$ |
| BN11 | AZ:BZ:CZ | $p(A Z)p(B Z)p(C Z)p(Z)$ | G15 (augmentation) | AZ:BZ:CZ | $p(A Z)p(B Z)p(C Z)p(Z)$ |
| BN12 | ABCZ _{A,B:C} | $p(Z ABC)p(A)p(B)p(C)$ | G1 | ABCZ | $p(Z ABC)p(ABC)$ |
| BN13 | ACZ _{A,C} :BZ | $p(Z AC)p(B Z)p(A)p(C)$ | G10 (augmentation) | ACZ:BZ | $p(Z AC)p(B Z)p(C A)p(A)$ |
| BN16 | BZ:CZ:A | $p(B Z)p(C Z)p(Z)p(A)$ | G17 (augmentation) | BZ:CZ:A | $p(B Z)p(C Z)p(Z)p(A)$ |
| BN17 | BCZ _{B,C} :A | $p(Z BC)p(B)p(C)p(A)$ | G7 | ABC:BCZ | $p(Z BC)p(B CA)p(A C)p(C)$ |
| BN19 | CZ:A:B | $p(Z C)p(C)p(A)p(B)$ | G10 | ABC:CZ | $p(Z C)p(ABC)$ |
| BN20 | A:B:C:Z | $p(Z)p(A)p(B)p(C)$ | G13 | ABC:Z | $p(Z)p(ABC)$ |

However, as pointed out above, the BN directed system lattice developed in this paper was constrained to disallow any DV that is a parent node within a V-structure. If this constraint were to be relaxed to allow DVs that are parent nodes in V-structures, then there are BN predictive models that give different analytical results than RA predictive models. Therefore, the BN directed system lattice developed in this paper is preliminary and incomplete.

To illustrate this point, consider BN17 from Figure 22 with its specific graph $ABZ_{A:B}:C$, and removing variable “C” for simplicity resulting in $ABZ_{A:B}$ with edge orientations $A \rightarrow Z \leftarrow B$ and with probability distribution $p(Z|AB)p(A)p(B)$. Here, Z is the DV and is the child node within the V-structure and is thus included within the BN directed system lattice developed in this paper. This graph is equivalent in terms of prediction to RA directed system graph G7 with specific graph $ABC:ABZ$. In contrast, consider BN17 with its specific graph $ABZ_{A:Z}:C$. Again, for simplicity and comparability, removing variable “C” results in $ABZ_{A:Z}$ with edge orientations $A \rightarrow B \leftarrow Z$ and with probability distribution $p(B|AZ)p(A)p(Z)$. Here, Z is the DV, and is a *parent* node within the V-structure; therefore, this specific graph was not considered in the BN directed system lattice developed in this paper. However, the predicting components within the probability distribution are different and thus will result in a different statistical result. The differences between $ABZ_{A:B}$ and $ABZ_{A:Z}$ are illustrated in Figure 39, in which a hypothetical joint probability distribution $p(ABZ)$, shown in (a), yields a

conditional distribution $p(Z|AB)$ for RA model ABZ and BN model $ABZ_{A:B}$, shown in (b), that is different from the conditional distribution $q(Z|AB)$ for BN model $ABZ_{A:Z}$, shown in (c). $ABZ_{A:Z}$ is an unconventional BN model in its choice of the parent node Z as the DV. These non-conventional BN models are not considered in this paper, but are a promising topic for future research that will extend the work reported here.

(a) $p(ABZ)$, joint distribution for ABZ

| | Z_0 | | Z_1 | |
|-------|-------|-------|-------|-------|
| | B_0 | B_1 | B_0 | B_1 |
| A_0 | 0.01 | 0.19 | 0.06 | 0.14 |
| A_1 | 0.17 | 0.31 | 0.03 | 0.10 |

(b) $p(Z|AB)$, conditional distribution for ABZ & $ABZ_{A:B}$

| | Z_0 | | Z_1 | |
|-------|-------|-------|-------|-------|
| | B_0 | B_1 | B_0 | B_1 |
| A_0 | 0.11 | 0.58 | 0.89 | 0.42 |
| A_1 | 0.83 | 0.75 | 0.17 | 0.25 |

(c) $q(Z|AB)$, conditional distribution for $ABZ_{A:Z}$
 where $q(ABZ) = p(B|AZ) p(A) p(Z)$
 and $q(Z|AB) = q(ABZ) / q(AB)$

| | Z_0 | | Z_1 | |
|-------|-------|-------|-------|-------|
| | B_0 | B_1 | B_0 | B_1 |
| A_0 | 0.21 | 0.74 | 0.79 | 0.26 |
| A_1 | 0.74 | 0.64 | 0.26 | 0.36 |

Figure 39 BN Directed System Prediction Example

Discussion

Neutral Systems

This paper builds on the RA work of Harris and Zwick (2021), which developed the BN neutral system general graph lattice of Figure 14, expanding it here to offer the BN neutral system specific graph lattice of Figure 20. This paper also builds on the joint RA-BN neutral system general graph lattice of Figure 35 developed in that earlier work, expanding it here to offer the joint RA-BN neutral system specific graph lattice of Figure 40. In developing these new lattices, this paper extends RA notation to encompass BN graphs (see the section titled “BN specific graph notation”).

For four variables, the joint RA-BN neutral system general graph lattice increases the number of general graphs from 20 in the RA lattice and 20 in the BN lattice to 30 in the joint RA-BN lattice, and unique specific graphs from 114 in the RA lattice and 185 in the BN lattice to 238 in the joint lattice. The integration of the two lattices offers a richer and more expansive way to model and represent complex systems leveraging the V-structure unique to BN graphs and the ability accommodate loops and hypergraphs in the RA lattice.

This paper also develops an algorithm to generate the joint RA-BN neutral system general and specific graph lattices for any number of variables in both upward and downward directions (section “Joint RA-BN algorithm”). The exhaustive and non-redundant RA and BN lattices follow the more general Rho lattice. Figure 40 shows the results of this algorithm for four variables. Although this algorithm is exhaustive, it does not create a hierarchical nesting of general or specific graphs. Such nesting is a desirable feature, so future extensions of this work could enhance the algorithm by enabling it to develop sequentially with each new graph being hierarchically nested. Given data, such an extension would allow statistical significance tests to be performed at each incremental step of lattice generation. Additionally, the current algorithm produces the exhaustive lattice, but searching the exhaustive lattice to find best candidate graphs is inefficient, so algorithms to efficiently search the joint lattice for best candidate graphs would be a useful extension.

Another promising extension of this work would be to develop hybrid RA-BN general graphs (Zwick, 2010) for neutral systems to further extend the expression of the joint RA-BN neutral system lattice developed in this paper. Such hybrid graphs could incorporate directed edges to encode BN V-structures with loops and hypergraphs found in RA. Other possible extensions of this work could explore the application of Bayesian networks to hypergraphs (Javidian, 2020) and under appropriate conditions to certain types of cycles (Forre, 2019).

Directed Systems

This paper develops the RA augmented directed system lattice (Figure 13), which is an extension of the conventional RA directed system lattice (Figure 11). While the conventional RA directed system lattice encompasses all prediction graphs in the BN directed system lattice (under the restriction that DVs in BN models are not parent variables in V-structures), the RA conventional directed system lattice did not include naïve Bayes graphs. Doing so, as shown in Figure 13, increases the number of general graphs from nine in the conventional RA lattice to 12 in the augmented lattice, and the number of specific graphs from 19 to 31. The augmented RA directed system lattice thus offers more candidate graphs, and this allows for the possibility of more accurate or simpler and thus more generalizable RA prediction models. Augmentation of the conventional RA directed system lattice was inspired in part by the BN directed system lattice developed in this paper.

Future extension of this work could examine whether BN graphs with predictions equivalent to RA models but with fewer degrees of freedom than RA predictive equivalents (because of independence constraints among the IVs) offer any advantage in calculations of statistical significance. If so, such BN graphs might replace their RA equivalents in the augmented directed system RA lattice. A related statistical issue that should be explored is how to compare augmenting directed RA models whose natural reference is A:B:..., the neutral system independence reference, with conventional directed systems models whose natural reference is AB...:Z, i.e., a reference that has an IV component that joins together all IVs in a single relation.

This paper develops the BN directed system lattice of prediction graphs for four variables (Figure 22), reducing the number of possible specific graphs from 185 in the BN neutral system lattice to 18 in the BN directed system lattice—a significant compression of the BN neutral system lattice when prediction of a single DV is the goal. This paper also shows that all of the graphs in the BN directed system lattice (where this lattice disallows graphs where the DV is a V-structure parent) are equivalent in their predictions to RA graphs, although many of them have fewer degrees freedom than their RA-equivalent counterpart. The augmented RA directed system lattice thus encompasses all of the BN directed system general graphs in terms of prediction, and offers additional predicative graphs, those including loops, that are not in the BN lattice. However, the restriction that disallows BN graphs where the DV is a V-structure parent might be relaxed, so a future extension of this work could consider expanding the BN

directed system lattice to include such unusual BN predictive graphs. An additional extension could be to develop an algorithm to generate the BN directed system lattice of general and specific graphs for any number of variables allowing for efficient search of the BN lattice for graphs that uniquely predict a single DV.

Joint RA-BN Four Variable Lattice of General and Specific Graphs

30 General Graphs, 238 Specific Graphs

Lattice of Four Variable General and Specific Bayesian Network Graphs

20 General Graphs, 185 Specific Graphs

Lattice of Four Variable RA General and Specific Graphs

20 General Graphs, 114 Specific Graphs

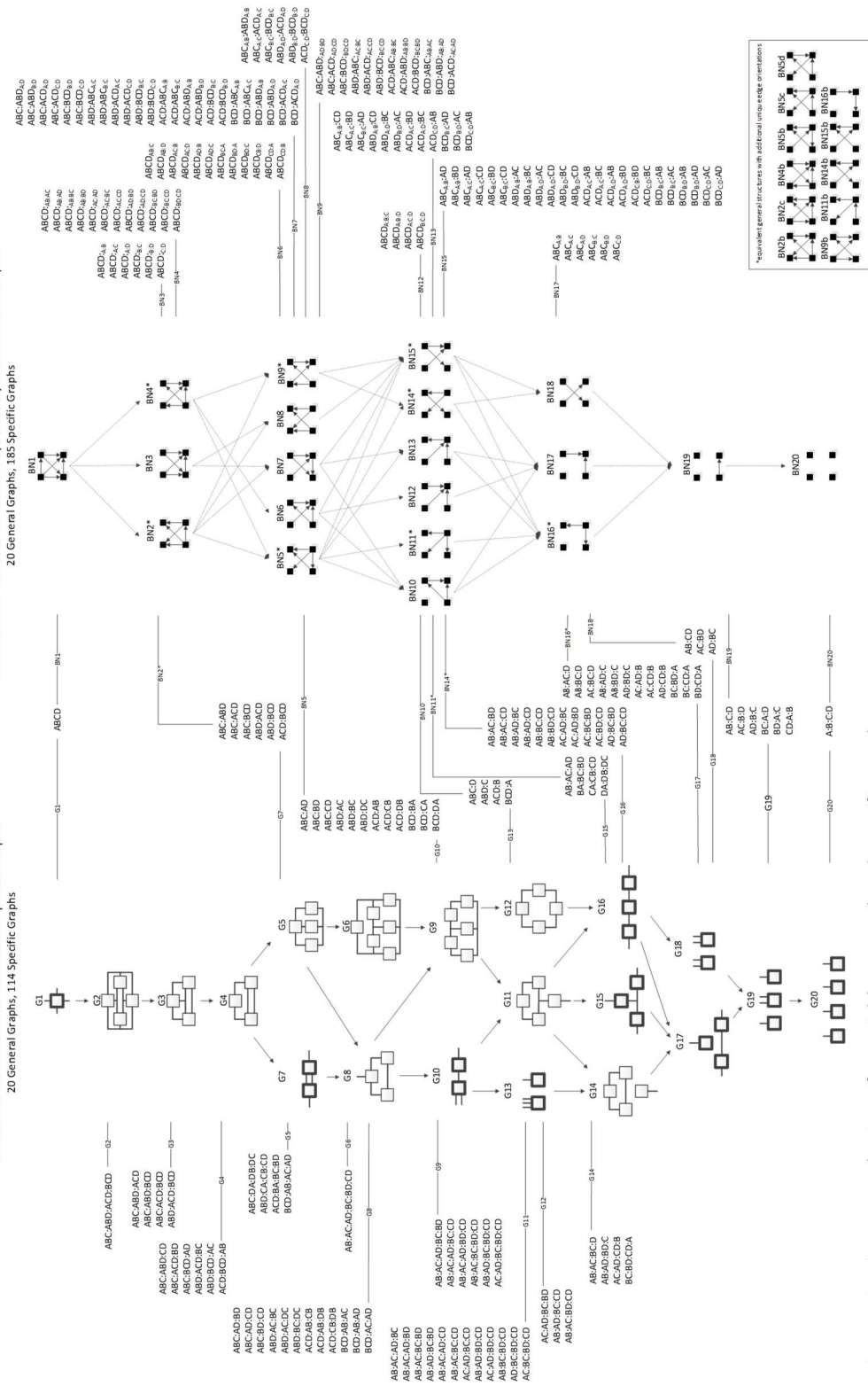


Figure 40 Joint RA-BN lattice of 4 variable general and specific graphs

Chapter 3

Paper 3 - Machine Learning Predictions of Electricity Capacity

Harris M, Kirby E, Agrawal A, Pokharel R, Puyleart F, Zwick M. Machine Learning Predictions of Electricity Capacity. *Energies*. 2023; 16(1):187.

Co-authored by Marcus Harris, Elizabeth Kirby, Dr. Ameeta Agrawal, Rhitabrat Pokharel, Frank Puyleart, and Dr. Martin Zwick.

Author contribution statement: Marcus Harris performed the conceptualization, formal analysis, writing original draft, and visualization. Elizabeth Kirby and Frank Puyleart performed conceptualization, writing on wind speed and wind power output, and writing review and editing. Dr. Ameeta Agrawal and Rhitabrat Pokharel performed calculations of neural network and support vector regression models and writing on neural network and support vector regression methods. Dr. Martin Zwick performed conceptualization, writing review and editing, and supervision.

DOI: <https://doi.org/10.3390/en16010187>

Link to published paper: <https://archives.pdx.edu/ds/psu/38933>

Abstract

This research applies machine learning methods to build predictive models of Net Load Imbalance for the Resource Sufficiency Flexible Ramping Requirement in the Western Energy Imbalance Market. Several methods are used in this research, including Reconstructability Analysis, developed in the systems community, and more well-known methods such as Bayesian Networks, Support Vector Regression, and Neural Networks. The aims of the research are to identify predictive variables and obtain a new stand-alone model that improves prediction accuracy and reduces the INC (ability to increase generation) and DEC (ability to decrease generation) Resource Sufficiency Requirements for Western Energy Imbalance Market participants. This research accomplishes these aims. The models built in this paper identify wind forecast, sunrise/sunset and the hour of day as primary predictors of net load imbalance, among other variables, and show that the average size of the INC and DEC capacity requirements can be reduced by over 25% with the margin of error currently used in the industry while also significantly improving closeness and exceedance metrics. The reduction in INC and DEC capacity requirements would yield an approximate cost savings of \$4 million annually for one of nineteen Western Energy Imbalance market participants. Reconstructability Analysis performs the best among the machine learning methods tested.

Introduction

The California Independent System Operator (CAISO) Western Energy Imbalance Market (WEIM) is an intra-hour electricity market with a footprint across much of the

Western US, including participants in California, the Pacific Northwest, the Southwest, and other areas. The market aims to balance supply and demand of electricity across this footprint on a 5-minute basis in the most economical way while satisfying a variety of constraints (CAISO, 2022a). Each participant must enter the market each hour with sufficient energy for its anticipated electricity demand (commonly referred to as “load”) and also with capacity to meet some amount of uncertainty in supply and demand. Prior research has shown that variable generation like wind and solar, and load forecasts (Ghosal, 2022; Jost, 2015; Lu, 2010; Obando-Ceron, 2018) are primary drivers of supply and demand uncertainty. The focus of this paper is on the appropriate amount of uncertainty that the CAISO should plan for each WEIM participant; more precisely, on the difference between actual supply and demand in a given time period that a WEIM participant should be able to compensate for with its INC and DEC capacities without assistance from other WEIM participants. This difference between supply and demand (commonly referred to as “imbalance”) is the dependent variable (DV, observed net load imbalance) being predicted in this study. (This DV is described section titled “Data”).

In order to establish the pool of generators that can be adjusted up or down, participants bid into the market INC and DEC capacity on an hourly basis to indicate the amount they are willing to adjust energy output up or down in the coming hour (along with a price for such adjustments). INC capacity is the capacity to respond to situations when demand exceeds supply. It is the ability to increase energy output within a certain

timeframe and is deployed when loads come in higher than expected and/or generators produce lower energy than expected. Participants called upon to increase generation are compensated for this deployment of energy. DEC capacity is the inverse, the capacity to respond to situations when supply exceeds demand. Specifically, it is the ability of generators to decrease energy output within a certain timeframe, called upon by the market operator and deployed by the participant when loads come in lower than expected and/or generators produce more energy than expected. Participants called upon to decrease generation save money because their sale to serve load is served by a cheaper market resource.

To participate fully in the WEIM for each hour, WEIM participants must meet four separate Resource Sufficiency Tests (FERC, 2022). Meeting these requirements provides the entire market with access to a large enough pool of adjustable generation and guarantees that participants are not relying on other market participants for their INC and DEC capacity needs. This pool of adjustable capacity ensures that generation supply can reliably meet load for all market participants in every time interval.

The Flexible Ramp Sufficiency Test (FRST) is one of the four Resource Sufficiency Tests (FERC, 2022; CAISO 2022b). This test establishes the minimum amount of INC and DEC reserves a participant must bid into the market in order to participate. Participants can bid more if they choose. This test is the focus of the analysis in this paper. The FRST, specifically the uncertainty component of the FRST, currently requires a participant to bid an amount of DEC capacity equal to the 2.5 percentile of “observed net load

imbalance” (which is Variable 1, the dependent variable, defined further in the section titled “Data”) for that participant, for the given hour, over the prior 40 weekdays, and to bid an amount of INC capacity equal to the 97.5 percentile over the same period. For weekends, the 2.5 and 97.5 percentiles are determined for the given hour, from the prior 20 weekend days (CAISO, 2022b). Setting the threshold at this level assumes that recent historical observations of net load imbalance are a good predictor of the potential range of imbalance in a coming hour. The goal of this approach is to create an INC and DEC upper and lower bound to ensure each participant is bringing enough INC and DEC capacity to support their historical net load imbalance. This approach is referred to as the Industry Model throughout the paper.

Prior research has shown that probabilistic methods like Bayesian Networks (Lu, 2010) and Monte Carlo simulation (Ghosal, 2022) improved the accuracy of net load imbalance forecasts. Further, prior research has shown different ML methods have advantages and disadvantages in predicting wind energy, solar energy and load forecasts (Sheraz, 2021).

Reconstructability Analysis (RA), Bayesian Networks (BN), Support Vector Regression (SVR), and Neural Networks (NN) are machine learning methodologies. RA and BN are both probabilistic graphical modeling methodologies and SVR and NN are other more commonly used machine learning methods. These four methods are applied in this paper to data provided by the Bonneville Power Administration to build point estimate prediction models of observed net load imbalance, the DV (point estimate

prediction models are described in the section titled “Results of DV prediction: comparing ML methods to Industry Model”). All of these methods underwent a 5 fold cross validation to ensure each model generalized well to new data. Midpoint predictions of the Industry Model and machine learning methods are compared in the section titled “Results of DV prediction: comparing ML methods to Industry Model” using three statistics: R squared, mean squared error, and mean absolute error.

The primary purpose of the first part of this study, where we used ML methods to build point estimate predictions of observed net load imbalance was to pick a good ML method for the second part of the study where, more importantly, INC and DEC predictions are made and compared to the Industry Model. That is, we do not here undertake a general comparison of the four ML methods, which would require applying them to multiple data sets; our aim in the first part of the study is primarily to identify a promising ML method for INC and DEC predictions.

The results of the point estimate predictions from the first part of the study show that three of the four methods do significantly better than the current Industry Model (CAISO, 2022b), and RA performed best overall; and the results of the second part of the study where RA is applied to make INC and DEC predictions show that the RA Model does significantly better than the Industry Model.

Metrics used to compare the RA INC and DEC prediction model to the Industry Model operationalize the general definitions published by the CAISO (CAISO, 2022c) and

are: Average Requirement (the average INC and DEC requirement over all observations in the dataset), Coverage (a measure that is the inverse of error, where error is how many times actual imbalance is outside the INC and DEC requirement), Closeness (how close the actual imbalance is to the INC and DEC requirement, whether greater or less than these requirements), and Exceedance (when, more specifically, imbalance is greater than the INC or DEC requirement, how much does it exceed the INC or DEC requirement). These metrics are described in detail in the section titled “Metrics for comparing the INC and DEC prediction efficacy”.

Of all the analyses in the paper, the results of the RA Model INC and DEC predictions have the greatest potential practical significance for the CAISO and WEIM participants. As shown in the test described in the section titled “Industry Model and RA Model INC and DEC prediction results,” the RA Model allows the FRST INC and DEC Average Requirement (the average amount of INC and DEC reserves that must be held) to be reduced by a total of 150.9 MW (To provide a sense of the scale of this reduction, 1 MW (one million watts) can power between 400 and 900 homes (NERC, 2012) depending on a number of factors), a 25.4% total reduction; specifically, 62.7 MW for INC reserves, a 23.0% reduction, and 88.2 MW DEC reserves, a 27.3% reduction while producing the same level of Coverage (the inverse of error) as the Industry Model. Additionally, Closeness and Exceedance metrics are also improved by the RA Model. These findings show that if the best RA model were used, the CAISO can retain the same Coverage as it has currently, while significantly lowering the INC and DEC requirement

for WEIM participants. Ultimately this has benefit to both the CAISO and WEIM participants in the form of lower cost. Conservatively, reducing the reserve requirement by this amount would result in approximately \$4 million (INC savings would be $\$168/\text{MW} - \text{day} \times 365 \text{ days} \times 62.7 \text{ MW} = \3.84 million . DEC savings would be $\$12/\text{MW} - \text{day} \times 365 \text{ days} \times 88.2 \text{ MW} = \386 thousand) in annual savings for the Bonneville Power Administration which is one of nineteen WEIM participants and which is the focus of the data analysis in this paper.

The following sections of this paper are organized as follows. The materials and methods section provides a description of the data used to perform the analysis and describes the four machine learning methods applied to the data as well as the Industry Model that is in current use. Results are then reported in two parts: an across-method comparisons of point estimate results and the RA Model INC and DEC predictions compared to the Industry Model. The last sections offer a discussion and conclusion focused on key findings and observations, and future possible research extensions.

Materials and Methods

Data

The data used in this analysis came from the Bonneville Power Administration which is a wholesale electric utility that began participation in the WEIM in the summer of 2022. The data is time series, in 15 min increments from January of 2014 to December of 2018. There are 172,175 observations, and there are no missing data.

There are 22 independent variables (IV) and one dependent variable (DV). 18 of the IVs are continuous, 4 are discrete. The DV is continuous.

Table 16 below lists the IVs and the DV, including characteristics about each variable and a short definition.

Table 16 Variable Names and Definitions

| Variable Number | Variable Name | Variable Name Single Letter Abbreviation | DV or IV | Continuous or Discrete | For RA/BN, number of bins | Time Lag | Short Description |
|-----------------|-----------------|--|----------|------------------------|---------------------------|-------------|--|
| 1 | NLFCErrorBoth_A | Z | DV | Continuous | 6 | 0 | Observed Net Load Imbalance. Positive values represent use of INC reserves to meet load requirements. Negative values represent use of DEC reserves to meet load requirements. |
| 2 | LoadRTD | A | IV | Continuous | 3 | 48 hour lag | Five-minute load measurement used in EIM 5-minute market optimization. |
| 3 | VERRTD | B | IV | Continuous | 3 | 48 hour lag | Five-minute VER measurement used in EIM 5-minute market optimization. |
| 4 | Net Load RTD | C | IV | Continuous | 3 | 48 hour lag | Five-minute measurement of load net VERs. Equal to LoadRTD - VERRTD. |
| 5 | NLMaxRTD | D | IV | Continuous | 3 | 48 hour lag | For each 15-minute interval, the maximum of the three 5-minute Net Load RTD values. |
| 6 | NLMinRTD | E | IV | Continuous | 3 | 48 hour lag | For each 15-minute interval, the minimum of the three 5-minute Net Load RTD values. |
| 7 | LoadFMM | F | IV | Continuous | 3 | 48 hour lag | Fifteen-minute load forecast used in EIM 15-minute market optimization. |

| | | | | | | | |
|----|--------------------------|---|----|------------|----|-------------|---|
| 8 | VERFMM | G | IV | Continuous | 3 | 48 hour lag | Fifteen-minute VER forecast used in EIM 15-minute market optimization. |
| 9 | NLFMM | H | IV | Continuous | 3 | 48 hour lag | Fifteen-minute forecast of load net VERs. Equal to LoadFMM - VERFMM. |
| 10 | VERFCFHfrtegt | I | IV | Continuous | 3 | 0 | Forecast of wind generation for the 15 minute time period the DV is being predicted, forecast made 48 hours prior |
| 11 | VERFCFHsvntwo | J | IV | Continuous | 3 | 0 | Forecast of wind generation for the 15 minute time period the DV is being predicted, forecast made 72 hours prior |
| 12 | VERFCFHninsx | K | IV | Continuous | 3 | 0 | Forecast of wind generation for the 15 minute time period the DV is being predicted, forecast made 96 hours prior |
| 13 | VERFCFHonetwty | L | IV | Continuous | 3 | 0 | Forecast of wind generation for the 15 minute time period the DV is being predicted, forecast made 120 hours prior |
| 14 | HLH/LLH | M | IV | Discrete | 2 | 0 | Heavy Load Hour - 6 a.m. to 10 p.m., Monday through Saturday, Light Load Hour - 10 p.m. to 6 a.m. Monday through Saturday and all day Sunday. |
| 15 | Season | N | IV | Discrete | 4 | 0 | Winter (Dec 1 - Feb. 28), Spring (Mar. 1 to May 31), Summer (June 1 - Aug. 31), Fall (Sept. 1 - Nov. 30) |
| 16 | Sunrise/Sunset | O | IV | Discrete | 2 | 0 | Sunrise (5am-7am), Sunset (5pm-7pm) |
| 17 | Hour of Day (2hr groups) | P | IV | Discrete | 12 | 0 | Hours of the day grouped into 2 hour groups, e.g. hours 1 and 2 grouped as 1, hours 3 and 4 grouped as 2, etc.) |
| 18 | NLFCErrBothx2x 1dayx | Q | IV | Continuous | 3 | 48 hour lag | 1 day average of the DV "NLFCErrBoth_A" |
| 19 | NLFCErrBothx2x 20dayx | R | IV | Continuous | 3 | 48 hour lag | 20 day average of the DV "NLFCErrBoth_A" |

| | | | | | | | |
|----|----------------------|---|----|------------|---|----------------|--|
| 20 | LoadRTD 24 Hr Avg | S | IV | Continuous | 3 | 48 hour lag | 24 hour average of IV "Load RTD" |
| 21 | LoadRTD24hrstdv | T | IV | Continuous | 3 | 48 hour lag | 24 hour standard deviation of IV "Load RTD" |
| 22 | LoadRTD 7 day Avg | U | IV | Continuous | 3 | 48 hour lag | 7 day average of IV "Load RTD" |
| 23 | LoadRTD7daystdv | V | IV | Continuous | 3 | 48 hour lag | 7 day standard deviation of IV "Load RTD" |

The DV (NLFCErrorsBoth_A which is synonymous with the term “observed net load imbalance”) is the difference between demand and supply (net load—generation), of electricity measured every 15 min. Positive values represent load in excess of generation and negative values represent generation in excess of load. The FRST establishes an INC and DEC range for each hour for each WEIM participant. For all the analysis in this paper, INC and DEC predictions described in the section titled “Results of INC/DEC prediction: comparing RA Model to Industry Model” and midpoint predictions described in the section titled “Results of DV prediction: comparing machine learning methods to Industry Model” are made in 15 min increments. That is, when any one of the machine learning methods makes a forecast for the following day, for a given hour, it is for a given 15 min increment within that hour, and the actual observed net load imbalance for that specific 15 min increment is compared to the prediction made for the same 15 min increment. For the Industry Model, the INC and DEC predictions and midpoint predictions for a given hour are the same for each of the four 15 min increments within that hour.

The IVs were selected based upon expert judgement gathered by surveying electrical engineers from the Bonneville Power Administration who are familiar with the variables that impact the DV, the Resource Sufficiency Tests, and WEIM operating requirements. In addition, in a survey of prior literature, Wind Forecast, Solar and Load Forecasts were shown to be primary predictors of net load imbalance (Ghosal, 2022; Jost, 2015; Lu, 2010; Obando-Ceron, 2018) and are included in the dataset in Table 16. The IV list of Table 16 is not exhaustive, as there may be other predictive variables that were not included, but it does provide a robust sampling of potentially predictive IVs.

Most IVs were lagged 48 h from the DV. Exceptions were time based IVs, for example the IV “Season” which represents one of four seasons (Winter, Spring, Summer, Fall). Another exception was “Wind Forecast” because the forecast was produced at least 48 h in advance. Lagging most variables was necessary in order to replicate the FRST INC and DEC prediction process (and the information available at the time the forecast is made) of the market operator (CAISO). The market operator needs to make the forecast (establishing the INC and DEC reserves requirement) at least 24 h in advance of the actual occurrence in order to be able to tell the WEIM participants what their minimum INC and DEC requirement is. A 48 h lag of the IVs was chosen in this study to be conservative, so that the data would definitely be available in time for the operator to use in the forecast establishing the INC and DEC requirement. Thus it is not unreasonable to expect that if the variables were lagged only 24 h, the machine learning method results would be even better than what is reported here.

Time based variables do not need to be lagged because they are known exactly. For example, what “Season” is 48 h ahead is known. The operator also knows, for example, what “hour of the day” is 48 h ahead. Similarly, for the Wind Forecast, the operator has a forecast of Wind Generation 48 h ahead, 72 h ahead, 96 h ahead, and 120 h ahead. All these forecast time frames are at, or greater, than the 48 h forecast the operator makes.

For both RA and BN, continuous data were binned into equal count bins, i.e., there were roughly an equal number of samples in each bin. For most IVs, three bins were used, i.e., low, medium, and high. By contrast, for SVR and NN, the continuous values were used. For all methodologies, the data was randomly divided into five equal folds (A, B, C, D, E) with 34,435 samples each. The original data had $N = 5 \times 34,435$ cases (172,175), each case separated from what preceded and what followed it by 15 min. There are no missing values for any of the IVs or the DV in the dataset. These folds were used for cross validation, and were used in the same way across all machine learning methods discussed in the following sections.

Methods

Four machine learning methods were applied to the data in addition to the standard Industry Model: RA, BN, SVR, and NN¹⁰. These methods were applied in order to produce the most accurate prediction of the DV possible that also generalizes well to

¹⁰ Input parameters for all methods are detailed in Appendix B.

withheld data, and these four methods were applied consistently, so that they could be compared for prediction efficacy to each other and to the Industry Model. For SVR and NN, we did not perform a hyper-parameter exploration as that would have been an entire study of its own, not warranted for the actual purpose of the first part of the study which was to use point estimate prediction success to select a machine learning method to model INC and DEC predictions in the second part of the study. Our use of RA and BN was similar in that we used the simplest form of RA and did not perform any preprocessing for RA or BN. We did not apply any elaborate pre-processing procedures in any of these four ML methods. Their results may thus be fairly compared.

Industry Model

The Industry Model establishes an upper and lower INC and DEC requirement based on the 2.5th and the 97.5th percentile of the DV over the prior 40 weekdays, for a given hour and for weekends and holidays over the prior 20 day weekend/holidays (CAISO, 2022b). The Industry Model INC and DEC predictions for a given hour are the same for all four 15 min increments within that hour.

Figure 41 shows an example of observed net load imbalance over a prior 40 day period, for a given hour of the day. This is an illustration of the type of data the Industry Model uses to establish the uncertainty component of the FRST, for the given hour, for the following day. The upper redline represents the INC requirement for the following day for the given hour, namely 400 MW. This is derived by taking the 97.5th percentile of observations over the prior 40 weekdays (circled in black) for the given hour.

Correspondingly, the DEC requirement is established the same way, the 2.5th percentile (circled in black) over the prior 40 weekdays establishes the DEC requirement, namely 500 MW. For weekends, the same procedure is applied, except the lookback is only 20 weekend days.

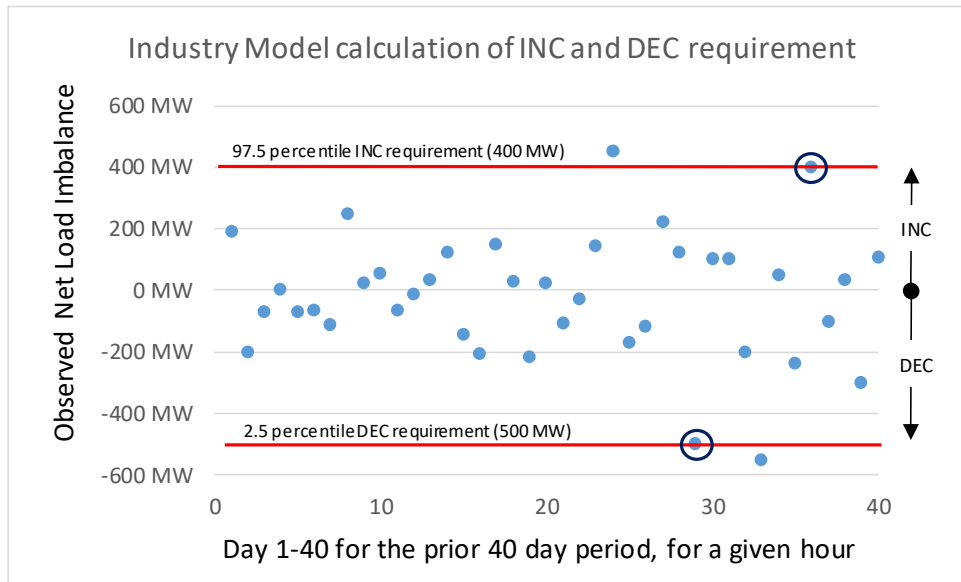


Figure 41 Industry Model calculation of INC and DEC requirement example

A midpoint prediction was derived from the Industry Model in order to compare to the machine learning (ML) methods assessed in this paper. The midpoint prediction for a given hour for the following day, is derived by taking the midpoint of the INC and DEC prediction. Using the example in Figure 41, the midpoint prediction would be -50 MW (midpoint of 400 MW INC and 500 MW DEC).

For the purpose of comparing the Industry Model to machine learning models, a two-step procedure is used. First, a point prediction is defined for the Industry Model that is the midpoint of the upper and lower INC and DEC thresholds, and the accuracy of

this point prediction is compared to the accuracy of the four machine learning point predictions. The results of this comparison are reported in the section titled “Results of DV prediction: comparing ML methods to Industry Model.” To be clear, the Industry Model does not on its own produce a midpoint prediction, it only produces an upper (INC) and lower (DEC) threshold prediction. Developing the midpoint prediction was necessary in order to compare the Industry Model to the midpoint predictions of the four other methods. In the second step, reported in the section titled “Results of INC/DEC prediction: comparing RA Model to Industry Model,” the Industry Model INC and DEC predictions are compared to the RA Model INC and DEC predictions, where the RA model was chosen over the other three machine learning methods because it gave the best point prediction results. The RA Model outperformed the Industry Model in both the first step of point prediction and the second step of INC and DEC prediction.

Reconstructability Analysis

RA is a data modeling approach developed in the systems community (Ashby, 1964; Broekstra, 1979; Cavallo, 1979; Conant, 1981, 1988; Klir, 1976, 1985, 1986; Krippendorff, 1979, 1981, 1986; Willett, 2004; Zwick, 2004a, 2004b, 2010, 2018) that combines graph theory and information theory. Its applications are diverse, including time-series analysis, classification, decomposition, compression, pattern recognition, prediction, control, and decision analysis (Zwick, 2001). RA is well suited for a problem in this domain, which is inherently probabilistic, and in which understanding the variables used to make predictions is important to system operators and WEIM

participants. RA and the other machine learning methods explicitly identify the independent variables (IVs) that are most predictive.

RA is designed especially for nominal variables, but continuous variables can be accommodated if their values are discretized. (RA could in principle accommodate continuous variables, but this extension of the methodology has yet to be formalized.) Graph theory specifies the structure of the model: if the relations between the variables are all dyadic (pairwise), the structure is a graph; if some relations have higher ordinality, the structure is a hypergraph. In speaking of RA, the word “graph” will henceforth include the possibility that the structure is a hypergraph. Graph structures are independent of the data except for necessary specification of variable cardinalities. In RA, information theory uses the data to characterize the precise nature and the strength of the relations. Data applied to a graph structure yields a probabilistic graphical model of the data (This paragraph from Harris, 2021).

RA can be applied to “neutral” and “directed” systems, and for both allows models that contain loops or do not contain loops. Neutral systems characterize the relations among all variables, and applications are common in network analysis and image processing. Directed systems characterize the relationship between IVs and the DV. (In principle, RA could accommodate multiple DVs, but the specific implementation of RA used in this study allows only a single DV.) For this application, RA directed systems are used because the primary goal is to predict the DV from the IVs. Further, for this analysis RA Models with no loops were used because the gain in using models with

loops was minimal and their computational cost was high. The computational time of models with loops is hyper-exponential with the number of variables whereas models without loops scale roughly polynomially with the sample size. Given the number of IVs under consideration, models without loops were the only models considered.

In RA, a lattice of structures is built given the number of variables under consideration (Zwick, 2004c). The lattice is typically searched from the bottom up, where the bottom is the independence model which is chosen as the reference. The goal is to find the model with the greatest amount of information content whose difference from this reference model is statistically significant. Alternatively, the model is searched downward from some starting model, typically the data, which is chosen as the reference model. In such downward search, the goal is to remove as many predictive relations as possible, thus reducing the complexity of the model, but where the model is statistically still not different from the data. For this application, a bottom up search was performed using the RA software OCCAM (Willett, 2004). In this upward search, a beam search¹¹ was used, where a search width of three was established as an input parameter. At the first level of search, the three most predictive models are retained, each using a single IV to predict the DV. So this first level identifies the three most predictive IVs. The second level of search begins with the results from the prior

¹¹ Not in the published paper: other algorithms to search the lattice could in principle be used, but Occam, the RA software, only implements a beam search.

level, and adds a best second predictive IV to each of the level one models, and so on to higher levels until a best model is found. Choice of the best model is done as follows.

As discussed previously, the data was divided into five equal folds. These 5 folds (A, B, C, D, E) were organized into five sets of training, validation, and test data as shown in Table 17. These fold names should not be confused with the single letter abbreviations of variables shown in Table 16.

Table 17 Cross Fold Validation

| Fold | Train | Validation | Test |
|-------------|--------------|-------------------|-------------|
| 1 | ABC | D | E |
| 2 | ABE | C | D |
| 3 | ADE | B | C |
| 4 | CDE | A | B |
| 5 | BCD | E | A |

For each of the five folds, the best RA Model was selected using the percent of the DV predicted correctly in the validation data. This best model, using parameters learned on training data only, was then applied to test data to compute the final results that are reported in the section titled “Results of DV prediction: comparing ML methods to Industry Model.”

As shown in the section titled “Data”, some of the data is discrete and some is continuous. For both RA and BN, continuous data were binned into equal count bins, i.e., there were roughly an equal number of samples in each bin. (By contrast, for SVR and NN, the continuous values were used.) The data contains no missing values, but for

more complex RA and BN models some IV states are not represented in the data or in some cases the IV state is not logically possible and therefore is absent from the data. This is common with very complex (high number of IVs and thus high degrees of freedom) RA or BN models. In the best models for the five folds, the number of missing IV states (Missing IV states defined as the percentage of individual IV states represented in the training data but not in the validation data) was on average 11%. To generate DV predictions for missing states, RA uses a simpler “backup model” chosen based on three criteria: 1. Highest average percent correct across all five folds based on validation data, 2. Its IVs are included in the best model for the fold, and 3. No missing IV states for the backup model. This resulted in a single backup model applied to all five folds; this is discussed further in Section titled “Best RA Model point estimate results.” The backup model is relevant only for making continuous predictions for missing IV states and for predicting the INC and DEC requirements; it was not used for selecting the best RA Model for the fold.

For the best RA Model, the B-Systems approach (Zwick, 2011) is applied to generate a continuous DV prediction for each IV state. This approach computes an expected value DV prediction from the conditional distribution of the model calculated (train data) given the IV state (Zwick, 2011), as follows:

DV prediction (IV_k) = $\sum_j p(DV_j|IV_k) rcv_j$, where, for the conditional probability $p(DV_j|IV_k)$, DV_j is j th bin of the DV, IV_k is k th bin of the IV, and where rcv_j is

a representative continuous value for each DV_j which is chosen to be the median continuous value for cases binned in this particular DV_j . The DV prediction result is thus the expected median DV value for a given IV state.

Bayesian Networks

BNs have origins in path model described in the 1930s (Wright, 1921, 1934) but it was not until the 1980s that BNs became more formally established (Neapolitan, 1989; Pearl, 1985, 1987, 1988). As does RA, BN combines graph theory and probability theory: graph theory provides the structure and probability theory characterizes the nature of relationships between variables. BNs are represented by a single type of graph structure; a directed acyclic graph, which is a subset of chain graphs, also known as block recursive models (Lauritzen, 1996). BNs can be represented more generally by partially directed acyclic graphs, a subset of chain graphs where edge directions are removed when directionality has no effect on the underlying independence structure. Discrete variables are most common in BNs, but BNs accommodate continuous variables without discretization (Driver, 1995) (As has been noted, in principle RA could also accommodate continuous variables but this feature has not yet been implemented.) For a three variable BN lattice, there are 5 general graphs and 11 specific graphs; for four variables there are 20 general graphs and 185 specific graphs with unique probability distributions (Harris, 2021). In the confirmatory mode, BNs can test the significance of a model relative to another model used as a reference (Tang, 2012); in the exploratory mode, BNs can search for the best possible model given a scoring metric. BNs are used

to model expert knowledge about uncertainty and causality (Neapolitan, 1989, Pearl, 1985) and are also used for exploratory data analysis with no use of expert knowledge (Rebane, 1987). Like RA, BN applications in machine learning and artificial intelligence are broad including classification, prediction, compression, pattern recognition, image processing, time-series, decision analysis and many others (This paragraph from Harris, 2021).

Augmented Naïve Bayesian Network

An augmented naïve Bayesian network (ABN) is an extension of the classic naïve Bayesian network classifier in which all IVs are independent of each other. ABNs relax the IV independence constraint to allow for IVs to be conditionally dependent upon each other given the DV. ABNs have been shown to produce better classification results than a naïve Bayesian network classifier (Friedman, 1997) and therefore were used in this paper as the BN prediction method.

The ABN algorithm uses parameters to restrict [or add] IV-IV edge connections based on how much they increase the maximum likelihood (or percent correct) results. For this analysis, standard input parameters were used. There was no restriction placed on the number of IV-IV edge connections and the prior link probability was set to 0.001. As discussed in the “Results” section, variations of the prior link probability do not produce materially better results. Varying the prior link probability increases the likelihood of adding or subtracting edge connections between IVs if the prior link probability is increased or decreased, respectively.

The data contains no missing values, however, for more complex BNs, there are some IV states that are not represented. For these missing states, BNs use the Expectation-Maximization (EM) algorithm (Dempster, 1977) to impute values for missing states. This differs from our implementation of RA, in which a simpler backup model, which is embedded in the best RA Model, is used to impute values for missing states as discussed previously. The EM algorithm is iterative, and more computationally costly than the purely analytic RA approach. This point is discussed in greater detail in the section titled “BN Comparison to RA Best Model.”

Genie Smile Software

The academic version of Genie Smile (BayesFusion, 2020) was the software used to apply ABNs to the data. An ABN search was applied to all 5 data folds. For BNs, in the same way as in RA, training data was used to define the model. The learned parameters on training data were used to compute results on the validation data. The best BN in terms of percent correct prediction accuracy on validation data was chosen as the best BN model. The best BN model, using parameters learned on training data only, was then applied to test data to compute the final results that are reported in the “Results” section.

To generate a continuous value BN prediction of the DV, the RA B-systems approach was applied in the same way using the representative continuous value of each DV bin multiplied by the model-calculated conditional probability of a that bin given the IV state.

Support Vector Regression

Support vector machines (SVM) were introduced in the late 1980s, but they came to prominence in 1995 (Vapnik, 1995). Over time, they became more and more popular with access to increasing computing power. Given a set of data points from two classes, in a two-dimensional plane, SVM draws a hyperplane that attempts to separate these data points into two categories. A margin is drawn on both sides of the hyperplane at an equal distance in such a way that the margin touches the closest point across both the classes of data points. The target of SVM is to maximize this distance between the hyperplane and the margin. Likewise, for SVR, there are data points on a two-dimensional plane, and a line is drawn such that it fits along the path of the data points and the points are as close to the hyperplane as possible. The margins around the hyperplane are drawn such that the points lying outside the margin are penalized. In short, the aim of SVR is to minimize the error (distance between the hyperplane and data points) for better generalization.

The data points are transferred to a higher dimension if there are more than two IVs.

Whether the data points are linearly separable or separable in any other way is determined by the function used which is called the kernel. Four different kernels were used in this analysis: radial basis function (rbf), linear, polynomial, and sigmoid.

Performance results are found in the section titled “Results of DV prediction: comparing ML methods to Industry Model.”

Neural Networks

The history of neural networks (NN) dates back to the late 1960s. Primitive neural networks were called perceptrons. Each perceptron resembles a biological neuron, i.e., it has inputs, a processing unit, and an output. The input values are assigned weights. The processing unit is a function that outputs a value after processing the input data. This function is called the activation function. These single-layer perceptrons could solve simple problems like “OR” gate and “AND” gate, but they could not solve an “XOR” gate problems. Later on, a backpropagation technique was introduced that updated the weights associated with the input after each iteration (Werbos, 1974). This technique led to the development of neural networks that are common today. Having many layers in between the input and output layers, an architecture called a multi-layer perceptron (MLP), also known as neural networks, allows one to model the XOR gate. MLPs are the model used for analysis in this research. Specifically, we implemented an MLP regressor built with 2 hidden layers each with 100 neurons.

Before SVM or MLP models were analyzed, an IV selection technique (feature selection) was used to select the k best IVs from the 22 IVs included in the dataset. The features were normalized using the MinMaxScaler preprocessing module from Sklearn to scale the values between 0 and 1. Then, the transformed data was fed to the f-regression model that calculates the correlation scores for each feature: the higher the score, the better the association between the IV and the DV.

Table 18 shows the 12 best features (i.e., $k = 12$, listed in alphabetical order) that were selected to be included in the SVR and MLP models. Out of 22 possible IVs, $k = 12$ was found to perform best for SVR and MLP on validation data and the list in Table 18 shows the 12 best that were used. These selected features were subjected to a second normalization, using the Sklearn MinMaxScaler.

Table 18 $k = 12$ features used in SVR and NN models

| | |
|----|-----------------------|
| 1 | HourofDayx2hrxgroups |
| 2 | Net Load RTD |
| 3 | NLFCErrorBothx2x1dayx |
| 4 | NLMaxRTD |
| 5 | NLMinRTD |
| 6 | Season |
| 7 | Sunrise/Sunset |
| 8 | VERFC_FH48fill |
| 9 | VERFC_FH72fill |
| 10 | VERFC_FH96fill |
| 11 | VERFC_FH120fill |
| 12 | VERFMM |

Both the SVR and MLP regressor models were implemented from the Sklearn python library (Pedregosa, 2011). All of the hyper-parameters were set to default values. These features are compared below to the predictors used in the RA and BN calculations.

Results of DV prediction: comparing ML methods to Industry Model

Best point estimate predictions and comparison of methods

The four machine learning methods and the Industry Model were assessed for their point estimate prediction efficacy based upon three performance statistics: R squared, Mean Average Error (MAE) and Mean Squared Error (MSE). For R squared, higher values are better; for MAE and MSE, lower values are better. The results in Table 19 show performance on test data from all five folds for the four prediction methods; for the Industry Model, the results are for all historical observations in folds A, B, C, D, E taken together.

Table 19 Results of Industry Model and Machine Learning Methods

| R Squared | | | | | | | |
|------------------|------------------|------------------|------------------|------------------|------------------|-----------|--------------------|
| Method | ABC Train E Test | ABE Train D Test | ADE Train C Test | CDE Train B Test | BCD Train A Test | Average | Standard Deviation |
| Industry Model | n/a | n/a | n/a | n/a | n/a | 7.5% | n/a |
| BN | 13.3% | 13.7% | 14.2% | 13.7% | 14.4% | 13.9% | 0.5% |
| RA | 33.5% | 33.2% | 35.2% | 33.2% | 34.1% | 33.8% | 0.9% |
| SVR-rbf | 7.5% | 7.5% | 7.5% | 7.2% | 8.0% | 7.5% | 0.3% |
| SVR-Linear | 6.3% | 6.4% | 6.5% | 6.1% | 6.9% | 6.4% | 0.3% |
| SVR-poly | 6.6% | 6.7% | 6.8% | 6.3% | 7.1% | 6.7% | 0.3% |
| SVR-sigmoid | 0.4% | 0.1% | 0.1% | 0.4% | 0.4% | 0.3% | 0.2% |
| MLP | 16.8% | 18.2% | 17.9% | 18.2% | 19.3% | 18.1% | 0.9% |
| MAE | | | | | | | |
| Method | ABC Train E Test | ABE Train D Test | ADE Train C Test | CDE Train B Test | BCD Train A Test | Average | Standard Deviation |
| Industry Model | n/a | n/a | n/a | n/a | n/a | 121.7 | n/a |
| BN | 103.0 | 102.2 | 102.4 | 103.4 | 102.7 | 102.7 | 0.5 |
| RA | 86.6 | 86.7 | 85.8 | 87.6 | 86.8 | 86.7 | 0.6 |
| SVR-rbf | 108.4 | 107.9 | 108.3 | 109.2 | 108.6 | 108.5 | 0.5 |
| SVR-Linear | 109.6 | 109.0 | 109.4 | 110.3 | 109.7 | 109.6 | 0.5 |
| SVR-poly | 109.1 | 108.6 | 109.0 | 109.9 | 109.4 | 109.2 | 0.5 |
| SVR-sigmoid | 588.3 | 579.6 | 580.7 | 600.5 | 582.8 | 586.4 | 8.5 |
| MLP | 100.5 | 99.2 | 99.8 | 100.4 | 99.7 | 99.9 | 0.5 |
| MSE | | | | | | | |
| Method | ABC Train E Test | ABE Train D Test | ADE Train C Test | CDE Train B Test | BCD Train A Test | Average | Standard Deviation |
| Industry Model | n/a | n/a | n/a | n/a | n/a | 27,339.7 | n/a |
| BN | 21,717.9 | 21,038.1 | 20,962.8 | 21,710.6 | 21,509.5 | 21,387.8 | 364.3 |
| RA | 16,717.4 | 16,425.5 | 15,894.2 | 16,904.0 | 16,616.8 | 16,511.6 | 386.0 |
| SVR-rbf | 23,164.5 | 22,576.3 | 22,603.6 | 23,361.5 | 23,164.7 | 22,974.1 | 359.9 |
| SVR-Linear | 23,470.0 | 22,822.8 | 22,860.1 | 23,631.9 | 23,410.9 | 23,239.2 | 372.2 |
| SVR-poly | 23,395.3 | 22,765.9 | 22,790.8 | 23,581.3 | 23,360.2 | 23,178.7 | 375.1 |
| SVR-sigmoid | 699,725.9 | 703,145.2 | 709,064.7 | 743,823.7 | 697,264.0 | 710,604.7 | 19,090.9 |
| MLP | 20,831.0 | 19,953.1 | 20,064.1 | 20,580.2 | 20,290.0 | 20,343.7 | 363.0 |

The results show that the RA method performed best on all measured statistics.

The first table within Table 19 shows that RA performed better than the Industry Model,

as follows. For the R squared statistic, the RA Model was on average 26.3% (33.8% – 7.5%) better than the Industry Model. For the MAE statistic, the RA Model was on average 35.0 better. For the MSE statistic, the RA Model was on average 10,828.4 better. The second to fourth tables within Table 19 compare RA to BN, the four types of SVR, and MLP, and these tables show that RA performed better than any of these other methods.

Note that our goal in this first analysis is not to accurately predict the DV time series, for which a low R squared would indeed be disappointing, but rather to pick the best machine learning method to use for the second part of the analysis described in the section “Results of INC/DEC prediction: comparing RA Model to Industry Model,” which is to analyze INC and DEC reserve requirements that must be held. In the first analysis we want to identify the ML method that best predicts the time series, but the precise accuracy of its predictions is not critical.

Best RA Model point estimate results

The RA Fold 5 model (BCD train, E validation, A test) performed best overall based on the percent correct statistic on validation data. Therefore, this model was chosen as the overall best RA Model (the best of the best models for all the folds). Table 20 shows results of training the best Fold 5 model on all 5 training sets, and statistics reported on all 5 test folds, including the use of a backup model where the overall best model did not offer predictions for some IV states. It is clear from the results that the model generalizes well to the withheld test data. The variance on all statistics reported

is very low across all five folds. Note that using the RA Fold 5 Model on the test data for all five folds slightly improves the three metrics: R Squared increases from 33.8% (Table 19) to 34.1% (Table 20); MAE decreases from 86.7 to 86.6; and MSE decreases from 16,511.6 to 16,451.8.

Table 20 Best RA Model tested on all folds

| Summary Statistic | ABC Train E Test | ABE Train D Test | ADE Train C Test | CDE Train B Test | BCD Train A Test | Average | Standard Deviation |
|------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|----------------|-------------------------------|
| R Squared | 34.0% | 33.7% | 34.9% | 33.5% | 34.1% | 34.1% | 0.5% |
| MAE | 86.3 | 86.4 | 86.1 | 87.2 | 86.8 | 86.6 | 0.4 |
| MSE | 16,591.6 | 16,266.2 | 15,981.2 | 16,803.5 | 16,616.8 | 16,451.8 | 326.5 |

Table 21 shows the individual IVs that resulted from the best model on all five folds. An RA Model throughout the paper is defined as the set of IVs used to predict the DV. For all five folds, the best model had 15 IVs. These are listed in Table 21, but it is important to realize that this list implies a 16-way interaction effect involving the 15 IVs and the DV. That is, if these 15 IVs were given single letter abbreviations from A to V from

Table 16 and if the DV were called Z, then the overall RA Best Model would be written as BFHIKLMOPQRSTUVWXYZ. Table 21 orders the variables to see where they are the same and different across all five folds. Of these 15 IVs, 12 (variables 4 through 15) were identical in the best models of all five folds. Only 3 IVs (variables 1 to 3) differed among these five best models. Bolded variables are the variables used in the backup model and are the same variables for all 5 folds. As noted and discussed below, an “overall best

model” was selected from these five best models; this overall best RA Model was the model for Fold 5.

Table 21 RA best model Independent Variables

| | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 Best Overall Model |
|----|---------------------------|----------------------------|----------------------------|----------------------------|---------------------------|
| 1 | NLMinRTD | NetLoadRTD | NLMinRTD | NetLoadRTD | VERRTD |
| 2 | LoadRTD | LoadFMM | LoadFMM | LoadRTD | LoadFMM |
| 3 | VERFMM | VERFMM | VERFMM | VERFMM | NLFMM |
| 4 | VERFCFHrtegt | VERFCFHrtegt | VERFCFHrtegt | VERFCFHrtegt | VERFCFHrtegt |
| 5 | VERFCFHninsx | VERFCFHninsx | VERFCFHninsx | VERFCFHninsx | VERFCFHninsx |
| 6 | VERFCFHonetwty | VERFCFHonetwty | VERFCFHonetwty | VERFCFHonetwty | VERFCFHonetwty |
| 7 | HLHLLH | HLHLLH | HLHLLH | HLHLLH | HLHLLH |
| 8 | SunriseSunset | SunriseSunset | SunriseSunset | SunriseSunset | SunriseSunset |
| 9 | HourofDayx2hrxgrou | HourofDayx2hrxgroup | HourofDayx2hrxgroup | HourofDayx2hrxgroup | HourofDayx2hrxg |
| 10 | NLFCErrorBothx2x1d | NLFCErrorBothx2x1da | NLFCErrorBothx2x1da | NLFCErrorBothx2x1da | NLFCErrorBothx2 |
| 11 | NLFCErrorBothx2x20 | NLFCErrorBothx2x20d | NLFCErrorBothx2x20d | NLFCErrorBothx2x20d | NLFCErrorBothx2 |
| 12 | LoadRTD24hravg | LoadRTD24hravg | LoadRTD24hravg | LoadRTD24hravg | LoadRTD24hravg |
| 13 | LoadRTD24hrstdv | LoadRTD24hrstdv | LoadRTD24hrstdv | LoadRTD24hrstdv | LoadRTD24hrstdv |
| 14 | LoadRTD7dayavg | LoadRTD7dayavg | LoadRTD7dayavg | LoadRTD7dayavg | LoadRTD7dayavg |
| 15 | LoadRTD7daystdv | LoadRTD7daystdv | LoadRTD7daystdv | LoadRTD7daystdv | LoadRTD7daystdv |

Of the variables that differed among the five folds, VERFMM is found in four of five folds. LoadFMM is found in three of five folds. NLMinRTD, LoadRTD, and NetLoadRTD are found in two of five folds. VERRTD and NLFMM are unique to fold 5. Even though variables 1–3 are not the same across all five folds, these variables are related to each other. For example, NLMinRTD and NetLoadRTD are variations of the same information; NLMinRTD is the minimum of NetLoadRTD; LoadRTD is the EIM 5 min load optimization forecast (for each 15 min time interval, there are 3, 5-minute load optimization forecasts, the value in the dataset is equal to the average of these 3 values); LoadFMM is the EIM 15 min load optimization forecast. Similarly, VERRTD is the

five minute EIM VER forecast, and VERFMM is the fifteen minute EIM VER forecast. In all folds, the model included one Net Load value (NetLoadRTD, NLMinRTD, or NLFMM), one Load value (LoadRTD, or LoadFMM), and one VER value (VERRTD or VERFMM).

Table 22 shows Fold 5 variables ranked in order of prediction efficacy of the DV using the information theoretic measure of percent reduction of uncertainty in the DV given the IV on training data. In the first table under “Joint Prediction Efficacy”, the ranking is based on the joint (cumulative) ability of the IVs to predict the DV. For example, if the model was allowed only one IV, it would be HoursofDayx2hrxgroups (Hour of the day). If the model was only allowed two IVs, HourofDayx2hrxgroups and VERFCFHrtegt (48 Hour Wind Forecast), ranked 1 and 2, respectively, would best predict the DV, reducing uncertainty in the DV by 7.47% jointly. If only allowed three IVs, HourofDayx2hrxgroups, VERFCFHrtegt, and SunriseSunset would best predict the DV, reducing uncertainty in the DV by 7.76% jointly, and so on. In the second table under “Individual Prediction Efficacy,” the IVs from the best fold 5 model are ranked based on their individual ability to predict the DV. For example, on its own, HourofDayx2hrxgroups reduces uncertainty in the DV by 5.36%, SunriseSunset reduces uncertainty in the DV by 1.98%, and so on. One can also see that, for example, variable 15 (LoadRTD7daystdv) individually only reduces uncertainty in the DV by 0.02%, however jointly in the first table, when combined with the prior 14 variables, reduces uncertainty in the DV by 2.18% (60.90–58.72%). Note that the variables in the backup

model (shown in bold) are not the first five individually predictive variables as may be expected due to the condition that the backup model must have 0% missing data.

Table 22 Rank order of variables for Fold 5 Best Overall Model

| Joint Prediction Efficacy | | | | Individual Prediction Efficacy | | |
|---------------------------|-----------------------------|---|---------------|--------------------------------|---|-------|
| Variable Name | Variable Name Single Letter | % Reduction Uncertainty in the DV Given IVs | Variable Name | Variable Name Single Letter | % reduction Uncertainty in the DV Given IVs | |
| 1 | HourOfDayx2hrxgroups | P | 5.36% | HourOfDayx2hrxgroups | P | 5.36% |
| 2 | VERFCFHfrtegt | I | 7.47% | SunriseSunset | O | 1.98% |
| 3 | SunriseSunset | O | 7.76% | VERFCFHfrtegt | I | 1.32% |
| 4 | HLHLLH | M | 8.01% | VERFCFHninsx | K | 1.05% |
| 5 | NLFCErrorBothx2x20day | R | 9.00% | VERFCFHonetwty | L | 0.73% |
| 6 | LoadRTD7dayavg | U | 10.52% | NLFCErrorBothx2x1dayx | Q | 0.51% |
| 7 | NLFCErrorBothx2x1day | Q | 13.48% | NLFCErrorBothx2x20dayx | R | 0.32% |
| 8 | VERRTD | B | 19.39% | HLHLLH | M | 0.30% |
| 9 | LoadRTD24hrstdv | T | 29.43% | LoadFMM | F | 0.22% |
| 10 | VERFCFHonetwty | L | 39.55% | NLFMM | H | 0.14% |
| 11 | LoadFMM | F | 46.27% | LoadRTD7dayavg | U | 0.09% |
| 12 | VERFCFHninsx | K | 51.39% | LoadRTD24hravg | S | 0.08% |
| 13 | NLFMM | H | 55.35% | LoadRTD7daystdv | V | 0.03% |
| 14 | LoadRTD24hravg | S | 58.72% | VERRTD | B | 0.03% |
| 15 | LoadRTD7daystdv | V | 60.90% | LoadRTD24hrstdv | T | 0.02% |

Best BN model point estimate results

Theoretically the BN method has the potential to perform as well as RA loopless models (Harris, 2021) which were used in this analysis. However, the ABN algorithm resulted in a simpler BN (less degrees of freedom) than the best RA Model. The prior link probability restrictions discussed in the section titled “Methods” result in a BN that is simpler in terms of model complexity and correspondingly give worse prediction results

than the RA Models. The best BN for each fold resulted in an average R squared of 13.9%, an MAE of 102.7 and an MSE of 21,387.8.

Figure 42 shows the best overall BN from Fold 5 which performed best across all five folds on the percent correct statistic. In this model, the IV LoadRTD7dayAvg (single letter notation: U) is the most interconnected IV, with seven IV-IV edge connections. In RA notation (Harris, 2021), using the single letter abbreviations of the variables listed in

Table 16, this best BN model is:

ABFENUSZ : BGUAERZ : CFDAERZ : DQHCZFZ : ERCQBPAZT : FDCAZ : GLBZ : HKDQZ : IJLZ : JIKOZ : KJHZ : LIGZ : MPUZ : NUVZ : OJSPZ : POEAMZ : QHDZ : RCBEZ : SOUAZ : TEVZ : UMSBANZ : VNUABTZ.

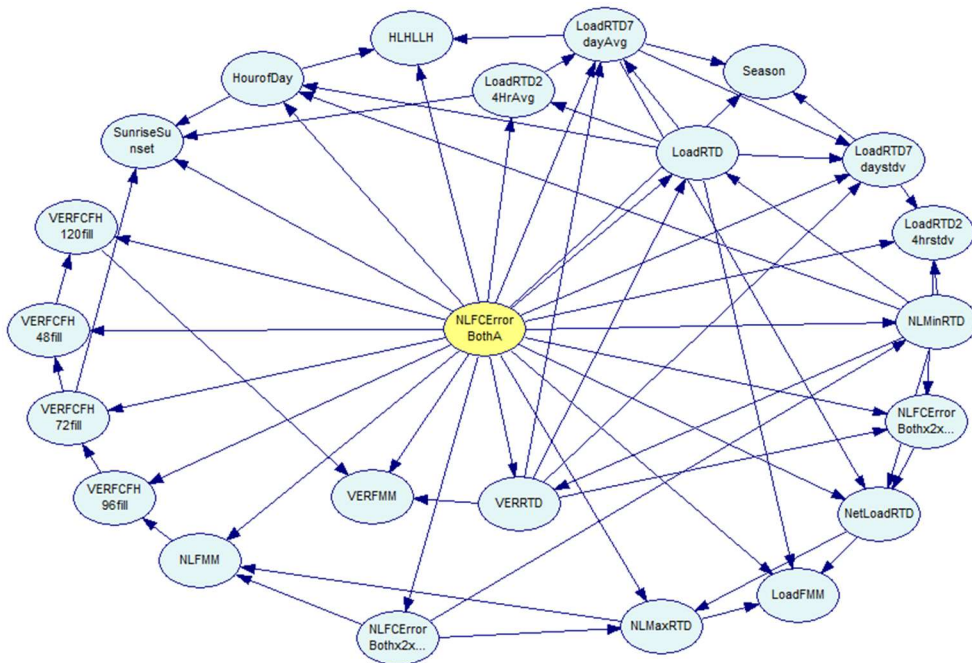


Figure 42 Best Overall BN

This includes all 22 IVs, which can be compared to the overall best RA Model, which includes only 15 IVs. That the RA model performed better than the BN model is, however, not actually surprising. In the RA Model, all 15 IVs are connected to one another and with the DV in a single very high ordinality predictive relation, while in this BN model there are 22 relations, of which the one with the most variables includes only 8 IVs, and most relations have many fewer IVs. Thus, while the best BN model includes all possible IVs as predictive, their interconnections are less extensive than variable interconnections in the RA Model; this lesser interconnectedness results in lower complexity in terms of degrees of freedom and lower prediction efficacy. Although the RA and BN methods have the potential to produce identical analytical results, the search algorithms specific to the particular Genie BN software produce significantly different results. These search algorithms implemented in the Genie BN software are, however, standard BN search algorithms.

A secondary test was performed to vary the prior link probability restriction from 0.001 to 0.002, 0.005, 0.01, and 0.0005, which has the effect of adding or removing edge connections between IVs, however this did not improve the models performance based on percent correct on validation data.

Table 23 shows the results of this test. Therefore, the standard link probability of 0.001 was used and resulted in the best BN model shown above.

Table 23 Test of different prior link probabilities

| | Link Probability | Percent Correct Validation |
|------------|-------------------------|-----------------------------------|
| Best Model | 0.0010 | 30.8% |
| Test 1 | 0.0020 | 30.3% |
| Test 2 | 0.0050 | 30.4% |
| Test 3 | 0.0100 | 30.2% |
| Test 4 | 0.0005 | 30.7% |

BN Comparison to RA Best Model

As a further supplementary analysis, the 15 IVs from the best RA Model were used to construct a BN that is similar to the best loopless RA Model. To do this, all IVs were connected to all other IVs, in addition to the standard IV-DV edge connection. This construct shown in Figure 43 is nearly analytically equivalent to the best Fold 5 RA Model. In RA notation (Harris, 2021), this BN model would be exactly the same as the overall best RA model, namely BFHIKLMOPQRSTUVWXYZ. The only difference is that the BN model uses the EM algorithm to impute missing values, or in this case missing IV states (there are no missing values in the raw dataset), whereas the RA Model uses a simpler backup model that is embedded in the best model to impute values for missing IV states. The prediction results of this show that RA and BN are nearly equivalent, as expected. The BN test was performed on the BCD training fold, and results on the A Test fold. For RA, the results are, 34.1% for R squared, 86.8 for MAE, and 16,616.8 for MSE. For the fully connected BN, the results are, 33.8% for R squared, 87.0 for MAE, and 16,719.7 for MSE. The RA and BN results are nearly identical, and for BN, these results

are far superior to the best ABN model found which had performance statistics of: 14.4% for R squared, 102.7 for MAE, and 21,509.5 for MSE. This analysis shows that BNs have the potential to produce models equivalent or nearly equivalent to RA. However, restrictions on IV interconnections, as well as significantly more computational cost for more complex BN models due to use of the EM algorithm, typically prevent the best BN model from allowing fully interconnected IVs which is standard and computationally very fast in loopless RA Models.

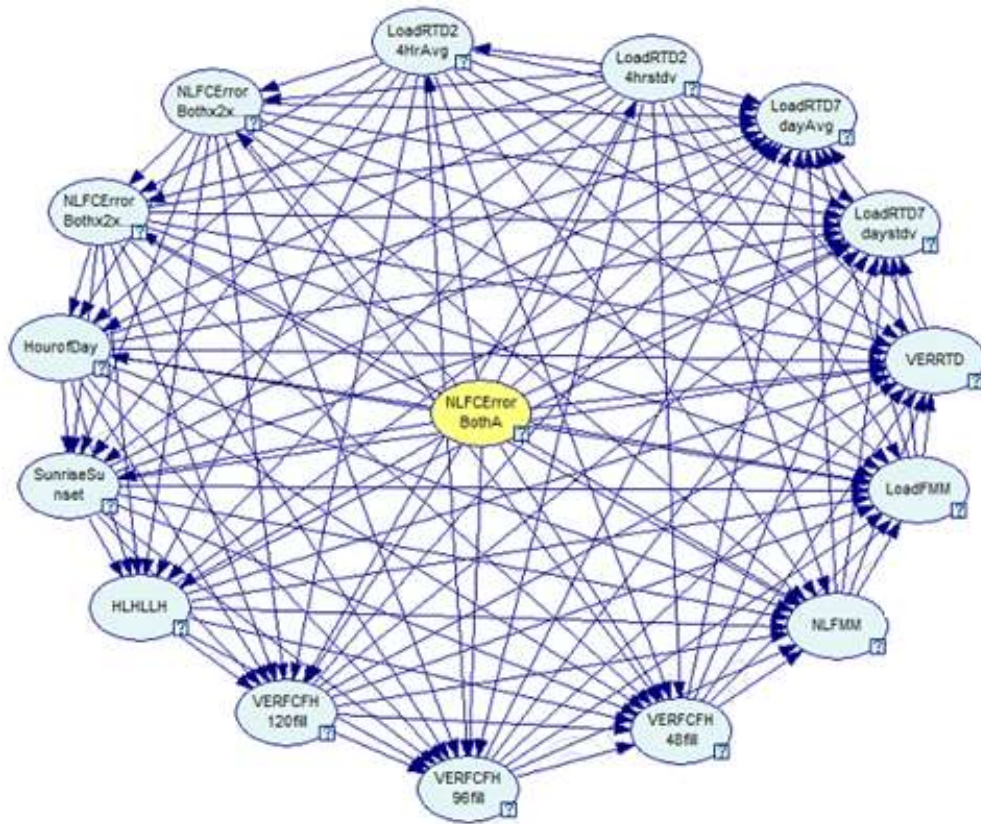


Figure 43 Best RA Model equivalent BN

To further illustrate this point, a simpler toy model was developed below to show that BNs with different edge orientations are equivalent in terms of prediction,

and that they are also fully equivalent to their RA counterpart. The theoretical basis for this concept was described extensively in prior work (Harris, 2021). This simpler toy example with only four variables was used to illustrate this feature of RA and BN because there are no missing states or missing values, thus the EM algorithm is not used, and a backup model also is not needed in RA to impute these missing values, so the calculated results for RA and BN are expected to be exactly equivalent. The results in Table 24 shows they indeed are equivalent.

Table 24 RA and BN Toy Example Results

| | R Squared | MAE | MSE |
|------------|------------------|------------|------------|
| RA ABC:ABZ | 12.935% | 103.738 | 21,842.710 |
| BN1 | 12.936% | 103.739 | 21,842.952 |
| BN2 | 12.936% | 103.739 | 21,842.952 |
| BN3 | 12.936% | 103.739 | 21,842.952 |
| BN4 | 12.936% | 103.739 | 21,842.952 |
| BN6 | 12.936% | 103.739 | 21,842.952 |
| BN12 | 12.936% | 103.739 | 21,842.952 |

Four variables were chosen for the toy example: NLFCErrBothA as the DV, HourOfDay, HLH/LLH, NLFCErrBothA2 × 1dayx as IVs. Figure 44 shows the RA specific graph G1 with these four variables and BN1, BN3, BN4, BN6 and BN12 described in prior work (Harris, 2021) and their corresponding network structures. All of these graphs have the same independence structure and thus, when data is applied, produce the same calculated conditional probability distributions and thus equivalent performance statistic results.

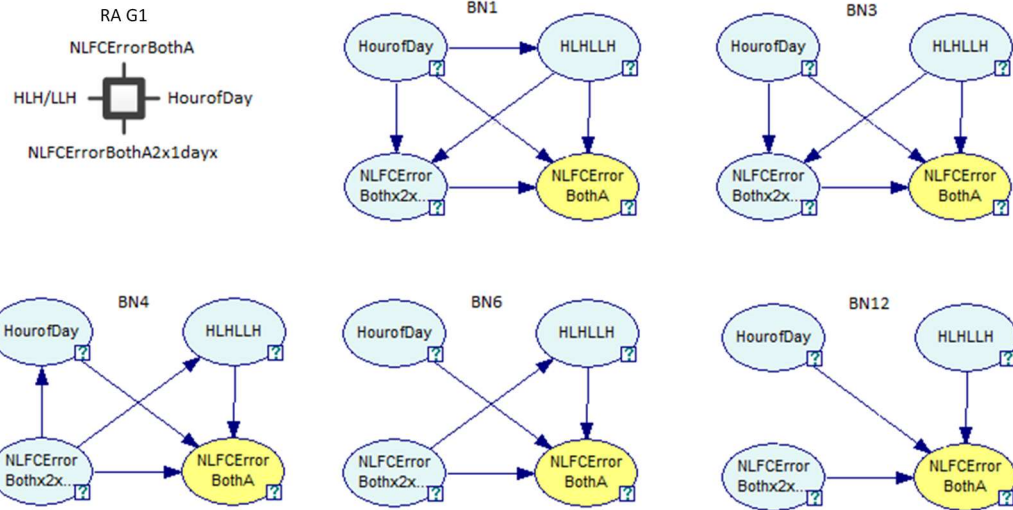


Figure 44 RA and BN Toy Examples

Table 24 shows the R squared, MAE, and MSE performance statistics for these graphs. Fold BCD was used to train each model, results shown are on validation Fold E. There is only a slight difference between the RA performance results and the BN performance results shown in Table 24. This is due to the OCCAM software using five significant digits for the RA calculations and the Genie Software that uses six significant digits.

SVM and NN point estimate results

As shown in the performance results in Table 19, among the four SVM kernels, rbf performed best, with an average R squared of 7.5% which is near identical to the Industry Model at 7.5%. MAE was 108.5 and MSE 22,974.1 compared to the Industry Model of 121.7 and 27,339.7, respectively. Rbf, although the best performing among SVR kernels, performed worse than RA, BN and MLP. MLP performed better than the Industry Model, BN, and SVR, but not as well as RA. MLP resulted in an average R

squared of 18.1%, MAE of 99.9 and MSE of 20,343.7 compared to the Industry Model of 7.5%, 121.7 and 27,339.7, respectively.

Comparing to the best RA Model which included 15 of 22 possible IVs, SVR and MLP performed best with 12 of 22 IVs as shown in Table 18. Of the 12, six were found in the best RA Model, four were also found in the RA backup model. Table 25 summarizes IVs found in the Best RA Model, SVR/MLP Model, and RA Backup Model, showing only those IVs that occur in at least two of these three models.

Table 25 IVs found in Best RA, NN/SVR, RA Backup Models

| | Fold 5 Best Overall RA Model | MLP/SVR | RA Backup Model |
|---|-------------------------------------|-----------------------|------------------------|
| 1 | VERFC_FH48fill | VERFC_FH48fill | VERFC_FH48fill |
| 2 | Sunrise/Sunset | Sunrise/Sunset | Sunrise/Sunset |
| 3 | HourofDayx2hrxgroups | HourofDayx2hrxgroups | HourofDayx2hrxgroups |
| 4 | NLFCErrorBothx2x1dayx | NLFCErrorBothx2x1dayx | NLFCErrorBothx2x1dayx |
| 5 | VERFC_FH96fill | VERFC_FH96fill | |
| 6 | VERFC_FH120fill | VERFC_FH120fill | |
| 7 | HLHLLH | | HLHLLH |

Results of INC/DEC prediction: comparing RA Model to Industry Model

RA INC and DEC prediction procedure

The previous section reports point estimate results for the four machine learning predictive methods. This section uses the RA Model, which performed best overall in point estimate predictions relative to the other machine learning methods, to generate INC and DEC predictions so that they can be compared to the results of the Industry Model INC and DEC predictions.

The Industry Model generates an upper INC requirement and the lower DEC requirement for a given future hour. As described previously, it uses the DV values over the prior 40 days, given the hour of the day, and establishes the INC requirement as the 97.5th percentile over the prior 40 day results and the DEC requirement as the 2.5th percentile over the prior 40 day results for weekdays. For weekends and holidays the DV is captured over the prior 20 weekend days. Although not a statistical prediction, this approach makes a persistence prediction (a prediction based on the prior history of the DV) about the future uncertainty of the DV. The Industry Model is not an analytical model that uses IVs to predict the DV; it is a purely empirical model that looks at past values of the DV.

As described in the “Best RA Model point estimate results” section, the overall best RA Model (from here forward called the “primary RA Model” to distinguish it from the backup model), is the best performing model among the machine learning methods tested for point estimate prediction efficacy. The midpoint (median) point estimate prediction of this primary RA Model for a given IV state is used as the reference point from which to generate an upper INC and lower DEC range prediction which is then tested against the performance of the INC and DEC threshold for each hour from the Industry Model.

However, the RA INC and DEC predictions were determined not from the primary RA Model itself, but rather from the backup model that was derived from (embedded in) this primary RA Model, its IVs being a subset of those in the primary

model. This is because the average frequency for individual IV states in the primary RA Model is too low for useful statistics. For the primary RA Model, the frequency of occurrence for a given IV state, averaged over all IV states, is 2.6 observations; for the backup model the average frequency is 425. The backup model offers a much higher frequency of observations for a given IV state, thus sampling an upper and lower threshold from this sample space is more robust.

From the primary RA Model, to generate the upper INC and lower DEC threshold predictions for a given IV state, the following RA INC and DEC prediction procedure was applied: For any given IV state,

1. Use the primary RA Model to predict the “primary” expected median value of the DV using the B-Systems procedure described in the section titled “Methods.”
2. Use the backup model (derived from the primary RA Model) to predict the expected median value for the DV using the B-Systems procedure.
3. Using the backup model, sample an upper and lower percentile (percentile amounts to be discussed in the paragraph below).
4. Subtract the lower percentile amount from the backup median found in step 2, and add it to the primary RA Model median found in step 1.
5. Subtract the upper percentile amount from the backup median found in step 2, and add it to the primary RA Model median found in step 1.

The following pseudocode is intended to further clarify the RA INC and DEC prediction procedure above.

Definitions:

- n = the number of IV states (all possible combinations of independent variable states)
- PEM = primary expected median = expected median DV value for an IV state (step 1 “primary”)
- BEM = backup expected median = expected median DV value for an IV state (step 2 “backup”)
- LPV = lower percentile value = lower percentile value for an IV state (step 3)
- UPV = upper percentile value = upper percentile value for an IV state (step 3)

Pseudo code for INC and DEC requirements for IV state j :

For IV state j from 1 to n .

$$\text{INC}(j) = \text{UPV}(j) - \text{BEM}(j) + \text{PEM}(j)$$

$$\text{DEC}(j) = \text{LPV}(j) - \text{BEM}(j) + \text{PEM}(j)$$

The result is a predicted upper INC and lower DEC threshold for each given IV state. The prediction is anchored on the median prediction for the IV state from the primary RA Model, but the range of uncertainty for a given IV state is generated from the backup model.

This procedure is represented in Figure 45. On the left of the graphic in Figure 45, “A”, for a given IV state, is the difference between the backup model median prediction for that IV state and the 97.5th percentile for the give IV state. “B” is the difference between the backup model median prediction and the 2.5th percentile. On the right hand side of the graphic, the INC prediction for a given IV state is the median prediction from the primary RA Model plus “A” and the DEC prediction for the same IV state is the primary RA Model median prediction less “B”.

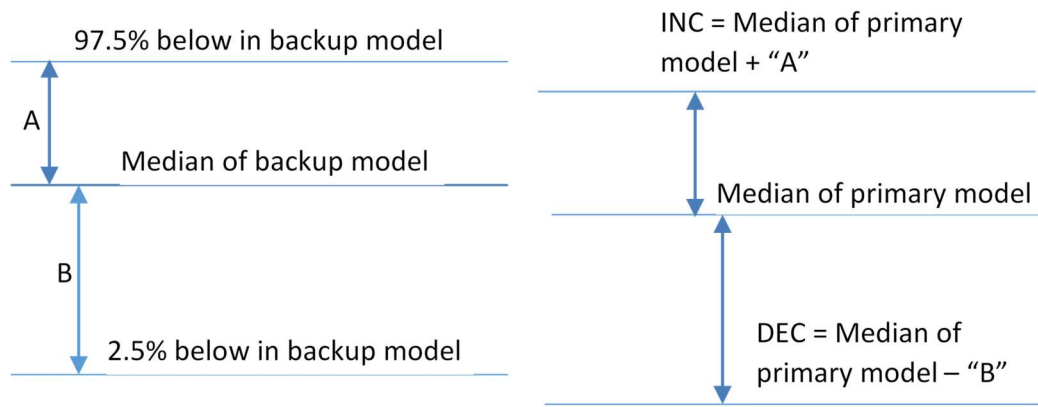


Figure 45 RA calculation of INC and DEC for a given IV state

The resulting model from this procedure will simply be called the “RA Model” for the remainder of the section.

Metrics for comparing INC and DEC predictions efficacy

The Industry Model is compared to the RA Model using four metrics the CAISO (CAISO, 2022c) uses to assess the efficacy of the Industry Model predictions of the FRST INC and DEC requirement. These four metrics are calculated and compared between the Industry Model and the RA Model. The four metrics are:

Average Requirement—The average INC requirement and average DEC requirement in MW over all observations in a given fold. Computed the same for the Industry Model and the RA Model. The higher the average requirement, the higher the cost of maintaining enough capacity to access the market. The equation for this metric is as follows:

$$1. \text{ Average INC Requirement} = \frac{1}{n} \sum_j \text{INC Requirement}_j$$

$$2. \text{ Average DEC Requirement} = \frac{1}{n} \sum_j \text{DEC Requirement}_j$$

where j is a given observation, n is the total number of observations in each fold (34,435).

Coverage—A measure of (the inverse of) error. Measured as the percentage of time that the observed net load imbalance falls within the model-produced INC and DEC requirement. The lower the coverage, the higher the frequency that the net load imbalance falls outside the INC and DEC requirement. The thresholds are set at 2.5% and 97.5% based on historical data, so the coverage aims at 95% or an error of 5%, but when applied to unseen data, the Industry Model is actually in error 7% of the time.

Closeness—The average difference in MW between the observed net load imbalance and the model-produced requirement when the observed net load imbalance falls either inside or outside the INC or DEC requirement. If the observed net load imbalance is positive, it is measured against the INC requirement; if negative, it is

measured against the DEC requirement. This metric measures how much the model is either over-estimating or under-estimating the capacity need, so the metric is actually an inverse of closeness. In this way, the metric can be thought of as Exceedance plus lost opportunity cost. The Closeness metric is reported separately for the INC and DEC requirements.

Exceedance—The average difference in MW between the observed net load imbalance and the model-produced requirement only when the observed net load imbalance falls outside the requirement, i.e., how much is the model under-estimating the need. The higher the exceedance value, the larger the gap between the INC or DEC capacity a participant contributed and what they actually needed, which can result in reliability issues if other market participants do not have unused capacity bid into the market. The Exceedance metric is also reported separately for the INC and DEC requirements.

For these metrics, one wants Average Requirement to be low, Coverage to be high, Closeness to be low, and Exceedance to be low.

To illustrate these metrics, Figure 46 provides 16 hypothetical examples of the INC and DEC requirements for the same time period of the RA Model and Industry Model, using the same 2.5%/97.5% thresholds, and of the (hypothetically) observed net load imbalance (the DV) for this time period. Each time period represents a different IV state for the RA Model and different historical data for the Industry Model; both the RA

Model and the Industry Model thus have different INC and DEC requirements in different time periods.

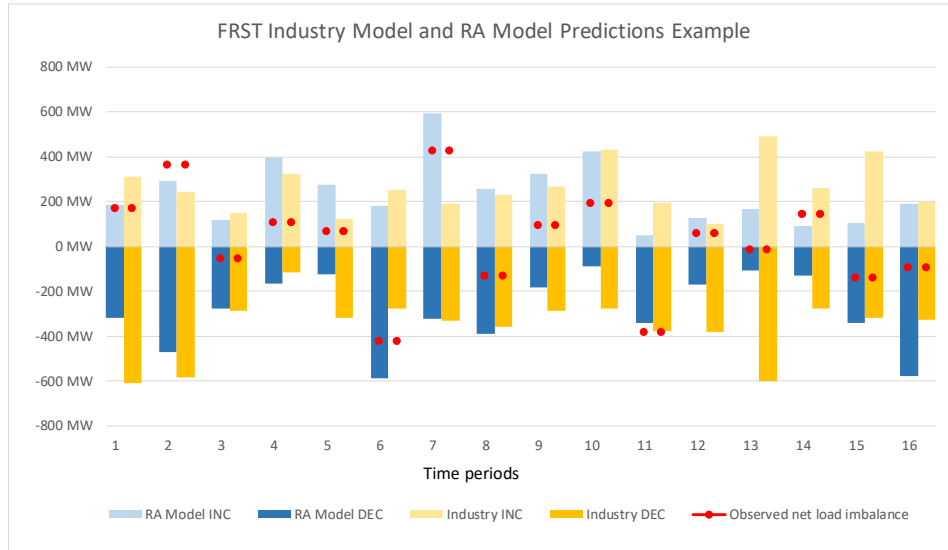


Figure 46 FRST Industry Model and RA Model Predictions Example

For example, in Figure 46, the Average Requirement metric for INC reserves is computed by taking the average of all the light blue bars for the RA Model and the average of all the light orange bars for the Industry Model.

Table 26 shows the data from Figure 46 in tabular form for ease of review.

Where bolded, the RA or Industry Model INC or DEC prediction is less than actual Observed Net Load Imbalance which results in an error and therefore a reduction in the Coverage metric. When not bolded, the RA or Industry Model INC or DEC prediction is greater than actual Observed Net Load Imbalance meaning the INC or DEC capacity requirement was sufficient to meet the Observed Net Load Imbalance in that interval.

Note that Figure 46 and the accompanying Table 26 are illustrative, and do not show the

advantages of the RA Model over the Industry Model; these advantages are clearly shown the Section titled “Industry Model and RA Model INC and DEC Prediction Results.”

Table 26 Observed Net Load Imbalance and INC and DEC predictions for each interval from Figure 46

| | | Observation number from Figure 46 | | | | | | | |
|-------|----------------|-----------------------------------|---------------|-------------|---------|--------|----------------|---------------|---------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | RA Model INC | 187 MW | 295 MW | 118 MW | 396 MW | 275 MW | 182 MW | 593 MW | 258 MW |
| 2 | Obs. Net Load | 171 MW | 363 MW | -57 MW | 106 MW | 65 MW | -422 MW | 425 MW | -132 MW |
| 3 | RA Model DEC | -318 | -472 MW | -276 | -165 MW | -124 | -586 MW | -324 MW | -388 MW |
| <hr/> | | | | | | | | | |
| 4 | Industry INC | 310 MW | 242 MW | 152 MW | 323 MW | 120 MW | 253 MW | 192 MW | 232 MW |
| 5 | Obs. Net Load | 171 MW | 363 MW | -57 MW | 106 MW | 65 MW | -422 MW | 425 MW | -132 MW |
| 6 | Industry DEC | -608 | -581 MW | -286 | -115 MW | -316 | -278 MW | -333 MW | -356 MW |
| <hr/> | | | | | | | | | |
| | | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 7 | RA Model INC | 326 MW | 423 MW | 51 MW | 125 MW | 166 MW | 93 MW | 104 MW | 188 MW |
| 8 | Obs. Net Load | 91 MW | 193 MW | -384 | 59 MW | -16 MW | 143 MW | -139 MW | -95 MW |
| 9 | RA Model DEC | -185 | -90 MW | -340 | -171 MW | -108 | -130 MW | -341 MW | -579 MW |
| <hr/> | | | | | | | | | |
| 10 | Industry Model | 268 MW | 434 MW | 192 MW | 101 MW | 493 MW | 261 MW | 421 MW | 197 MW |
| 11 | Obs. Net Load | 91 MW | 193 MW | -384 | 59 MW | -16 MW | 143 MW | -139 MW | -95 MW |
| 12 | Industry Model | -287 | -276 MW | -376 | -383 MW | -602 | -276 MW | -316 MW | -327 MW |

The Coverage metric is computed by summing the number of times the observed net load imbalance falls within the INC and DEC range over all time periods. Time period 7, in Figure 46 or Table 26, shows where the RA Model has Coverage and the Industry Model does not as the observed net load imbalance is within the RA INC Requirement but greater than the Industry Model INC requirement. For the 16 time periods in the figure, the RA Model has Coverage in 13 time periods, while the Industry Model has Coverage in 12 periods.

The Closeness metric for INC reserves is computed as the difference between the observed net load imbalance and the INC requirement when the observed net load imbalance is positive, and the average over all time periods. The Closeness metric is the same for DEC except it's applied against the DEC requirement when the observed net load imbalance is negative. Time period 14 in Figure 46 or Table 26 provides an example where the RA INC Closeness is slightly better than the Industry Model INC Closeness. However, this highlights the fact that Coverage is more important than Closeness because in this example even though the RA Model INC requirement is closer to the observed net load imbalance, the RA Model is in error whereas the Industry Model is not.

The Exceedance metric is the average difference between observed net load imbalance and the INC requirement when the observed net load imbalance is greater than the INC requirement. The same is computed for DEC when the observed net load imbalance is less than the DEC requirement. Time period 2, in Figure 46 or Table 26, shows an example where both the observed net load imbalance exceeds the RA Model INC requirement and the BN Model INC requirement but it is exceeded more for the Industry Model than the RA Model.

Table 27 shows summary statistics for the four metrics computed based on the 16 samples from Figure 46. For these metrics, one wants Average Requirement to be low, Coverage to be high, Closeness to be low, and Exceedance to be low.

Table 27 Calculated metrics from Figure 46 example

| | Industry Model | | | RA Model | | | Delta RA Less Industry | | |
|------------------|----------------|-------|-------|----------|-------|-------|------------------------|--------|--------|
| | INC | DEC | Total | INC | DEC | Total | INC | DEC | Total |
| Avg. Requirement | 261.9 | 357.1 | 619.1 | 236.2 | 287.3 | 523.5 | (25.7) | (69.8) | (95.6) |
| Coverage | 87.5% | 87.5% | 75.0% | 87.5% | 93.8% | 81.3% | 0.0% | 6.3% | 6.3% |
| Closeness | 149.2 | 228.8 | N/A | 148.1 | 208.9 | N/A | (1.1) | (19.8) | N/A |
| Exceedance | 177.1 | 75.9 | N/A | 59.3 | 43.3 | N/A | (117.9) | (32.6) | N/A |

Industry Model and RA Model INC and DEC predictions results

A test was performed on the RA Model to compare the prediction efficacy to that of the Industry Model. The Average Requirement, Coverage, Closeness, and Exceedance metrics are reported for the Industry Model and the RA Model and the delta between them.

The test scaled the RA Model upper and lower threshold percentiles so that the RA INC and DEC produced the exact same Coverage between the two models. For example, referring to Figure 46, this would have meant scaling the RA upper and lower thresholds so that of the 16 observations, there were the exact same number of INC errors and DEC errors between the Industry Model and the RA Model: when the observed net load imbalance was positive, it fell outside the INC requirement the same number of times in both the Industry Model and the RA Model, and when the observed net load balance was negative, it fell outside the DEC requirement the same number of times in both the Industry Model and the RA Model. This test is intended to see if after fixing Coverage of the RA Model to be identical to the Industry Model the Average

Requirement, Closeness and Exceedance metrics are reduced under the RA Model relative to the Industry Model.

Reduction in all three other metrics for the RA Model is indeed found. Table 28 shows the resulting statistics. As can be seen in Table 28, while Coverage is held the same for the Industry and RA Model, the Average Requirement, Closeness and Exceedance are all reduced relative to the Industry Model. This test shows that if the Industry Model Coverage (error rate) is acceptable, because the RA Model makes a more accurate prediction of the upper and lower INC and DEC thresholds, the Average Requirement for INC and DEC can be reduced. On average, 62.7 MW less INC, 88.2 MW less DEC, and 150.9 MW less total capacity would have to be held, while still maintaining the same Coverage. Further, Exceedance is also lower, 7.1 MW lower for INC, 12.3 MW lower for DEC. Even though Coverage is the same as the Industry Model, when the observed net load imbalance exceeds the RA Model INC or DEC requirement, it exceeds by less, on average, than the Industry Model.

Table 28 Final Results

| | Industry Model | | | RA Model | | | Delta RA Less Industry | | |
|------------------|----------------|-------|-------|----------|-------|-------|------------------------|--------|---------|
| | INC | DEC | Total | INC | DEC | Total | INC | DEC | Total |
| Avg. Requirement | 272.8 | 322.4 | 595.3 | 210.1 | 234.3 | 444.3 | (62.7) | (88.2) | (150.9) |
| Coverage | 95.4% | 97.6% | 93.0% | 95.4% | 97.6% | 93.0% | 0.0% | 0.0% | 0.0% |
| Closeness | 190.2 | 249.1 | N/A | 153.1 | 200.3 | N/A | (37.1) | (48.8) | N/A |
| Exceedance | 99.6 | 121.3 | N/A | 92.4 | 109.0 | N/A | (7.1) | (12.3) | N/A |

Figure 47 shows more detail about the results of Test 1 and the Exceedance metric. Given the same 93% Coverage (7% error) for the RA and Industry Model, when the RA Model (blue bars) is in error, the error is smaller than the Industry Model.

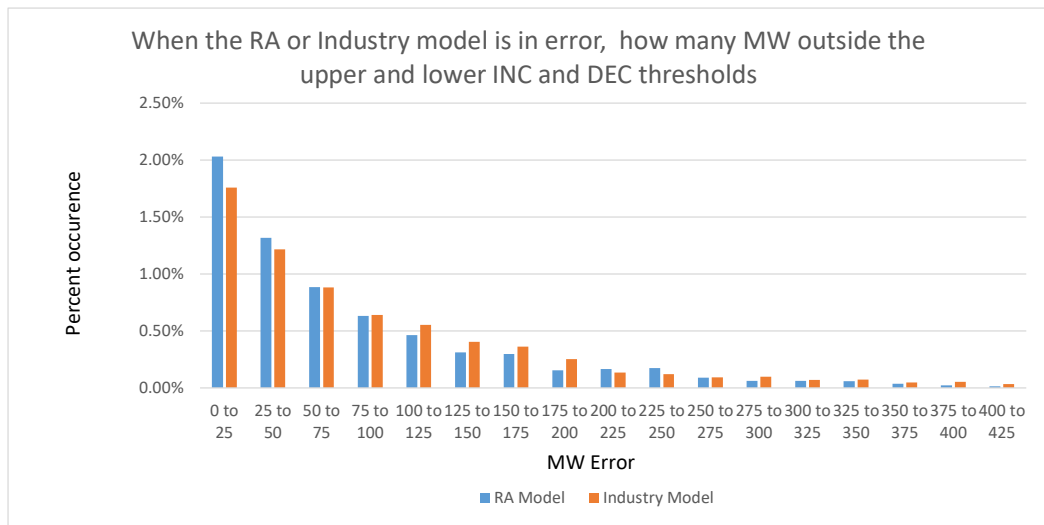


Figure 47 Test Results, Total Error and MW magnitude

RA Backup Model INC and DEC prediction results

The RA backup model was also tested by itself to show that the point estimate and using the RA INC and DEC prediction procedure is superior to using the backup model by itself. The model was tested on BCD training data and A test data. The backup model performed better than the Industry Model on the single point prediction, but

worse than the RA Model in terms of R squared, MAE, and MSE, as expected. Table 29 shows a comparison of the resulting summary statistics. On this fifth fold, the RA Backup Model also performed better on R squared, MAE, and MSE than the BN model and all SVR models, but not as well as the MLP model.

Table 29 RA Backup Model Point Estimate Summary Statistics

| | Industry Model | RA Model | Backup RA Model |
|-----------|-----------------------|-----------------|------------------------|
| R squared | 7.5% | 34.1% | 15.0% |
| MAE | 121.7 | 86.8 | 102.4 |
| MSE | 27,339.7 | 16,616.8 | 21,409.9 |

The backup model was also assessed for its ability to predict INC and DEC capacity using the same test previously applied to the RA Model. The test held Coverage equal to the Industry Model to see if the Average Requirement was reduced.

Table 30 shows how the backup model performed relative to the RA Model. The three numbers shown are deltas from the best RA Model. For the same Coverage, the backup model required 64.8 MW more Total capacity than the RA Model. This comparison shows that the backup model (where INC and DEC are applied to the backup model point prediction) did not perform as well as the Best RA Model (where the INC and DEC of the backup model are applied to the Best RA Model point prediction).

Table 30 RA Backup Model INC and DEC Predictions Compared to RA Model

| | Increase in Avg. Requirement |
|------------|-------------------------------------|
| Total Avg. | 64.8 |
| INC Avg. | 30.3 |
| DEC Avg. | 34.5 |

Discussion

The RA method performed better than the other machine learning methods applied in this research when comparing single point estimate predictions for a given observation (section titled “Results of DV prediction: comparing ML methods to Industry Model”). Our presumption is that RA performed better than the other methods because it is able to efficiently determine the optimal set of IVs to be used in the model whereas MLP and SVR used f-regression to select the predictive IVs, and the BN algorithm is limited in network complexity by the computational cost of the EM algorithm to make predictions about missing values or missing states. Although we did not perform a hyper-parameter exploration for SVR and NN as that would have been an entire study of its own, it was not warranted for the actual purpose of the first part of the study which was to use point estimate prediction success to select a machine learning method to model INC and DEC predictions in the second part of the study. Our use of RA and BN was similar in that we used the simplest form of RA and did not perform any preprocessing for RA or BN. Thus all four ML methods did not utilize an elaborate preprocessing and were thus treated equally.

The best RA Model resulted in 15 of the 22 possible IVs being included in the model. Of these, the most predictive variables are hour of day (HourOfDayx2hrxgroups), the 48 hour wind forecast (VERFCFHFrtegt), and Sunrise/Sunset (SunriseSunset) found in

Table 22. These variables were also found and included in the best SVM and MLP models. The Sunrise/Sunset IV is a time based IV that reflects 5 a.m.–7 a.m. PST and 5 p.m.–7 p.m. PST. It is a simple variable that was intended to capture some of the morning and evening solar and load ramp uncertainty.

Similar to the Sunrise/Sunset variable, the 48 hour wind forecast variable reflects times of energy imbalance uncertainty. Its information content is useful since knowing the wind forecast provides information about the maximum amount of INC that could be needed and maximum amount of DEC that could be needed. Additionally, the relationship between wind power output and wind speed shown in Figure 48 illustrates why the wind forecast is an important predictive variable. The greatest uncertainty in wind power output is when the wind forecast is in the middle of the nameplate (total possible output) of a plant. At low and high wind power output, the output changes little for small changes in wind speed, whereas at medium output, the output changes more significantly for the same small change in wind speed. Thus, actual wind power output is more variable when forecast output is around half of the plant nameplate. The RA Model is improved relative to the Industry Model by encompassing the wind forecast. The CAISO has been considering adjusting its methodology to account for the wind forecast, and the RA Model results and empirical results suggest that the

predictive information contained in the wind forecast has the potential to significantly improve INC and DEC predictions.

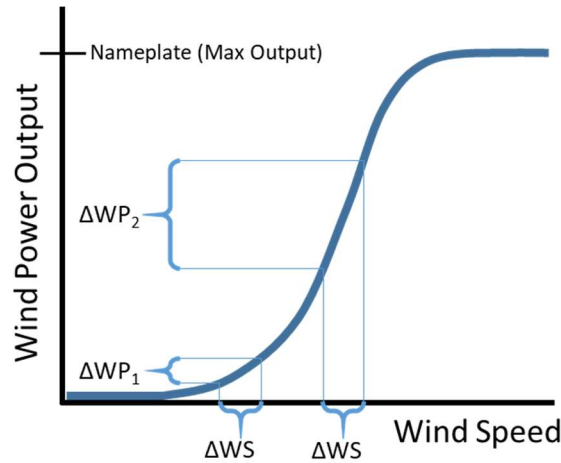


Figure 48 Relationship of Wind Power Output to Wind Speed

Additionally, interesting, the best RA Model, and each best model from all 5 folds shown in Table 21 included the 48 h wind forecast, 96 h wind forecast and 120 h wind forecast. This is an indication that there is additional, significant information, contained in the 96 and 120 h wind forecasts not contained in the 48 h wind forecast. This is supported by work previously done at Bonneville Power Administration. This work found that larger differences in wind power output forecasts of different time durations result in increased uncertainty in actual wind power output, while smaller differences in wind power output forecasts of different time durations result in reduced uncertainty in actual wind power output. Our presumption is that the differences between forecasts of wind power output of different time durations is a result of greater or less uncertainty in the weather pattern. While the CAISO has considered adjusting their model to account for the wind forecast, this research suggests that including the closest-in wind forecast is

useful and also including forecasts of longer time durations could further improve the prediction results.

Last, it was surprising that each of the best fold models from Table 21 contained an RTD variable, Load variable and FMM variable but never two of the same type of variables in a single model. For example, the best Fold 1 model contained NLMinRTD and Fold 2 contained NetLoad RTD, which are related but different IVs, but no best fold model included both of these Net Load variables, nor both Load or FMM variables.

Conclusions

The primary aim of this research was to build an INC and DEC capacity prediction model that improves upon the current Industry Model. The results in this paper (section titled “Results of INC/DEC prediction: comparing RA Model to Industry Model”) show that the best RA Model does in fact do that. As shown in the test results, the best RA Model reduces Total reserves relative to the Industry Model by 150.9 MW on average, a reduction of 25.4%; INC reserves are reduced on average 62.7 MW, a reduction of 23.0% relative to the Industry Model, while DEC reserves are reduced on average 88.2 MW, a reduction of 27.3. Additionally, Closeness and Exceedance metrics are also reduced significantly relative to the Industry Model. Conservatively, reducing the reserve requirement by this amount would result in approximately \$4 million (INC savings would be $\$168/\text{MW-day} \times 365 \text{ days} \times 62.7 \text{ MW} = \3.84 million . DEC savings would be $\$12/\text{MW-day} \times 365 \text{ days} \times 88.2 \text{ MW} = \386 thousand) in annual savings for the

Bonneville Power Administration which is one of nineteen WEIM participants and which is the focus of the data analysis in this paper.

This paper also discusses the results and theoretical comparisons between RA and BN as they are both machine learning methods that are analytically very similar. This research built on prior work (Harris, 2021) and used a concrete example to show that the best RA Model identified in this research can be replicated exactly in a BN. However, the BN search algorithm, in particular the computational cost of the EM algorithm, prevent such complex BN models from being found by standard BN search algorithms. Further, this research (Figure 44) used toy examples to show that several BNs with different edge topologies are analytically equivalent to a single RA graph, these toy examples further amplify and illustrate prior work (Harris, 2021).

Another aim of this research was to identify predictive variables of net load imbalance. This research also accomplishes this aim showing that wind forecast, sunrise/sunset, and hour of day are primary predictors of net load imbalance and should be considered for inclusion in any future industry application.

Logical extensions of the methods comparison could test other pre-processing steps and hyper-parameter exploration for MLP and SVR such as z-score (Buzau, 2018), clustering evaluation criteria (Buzau, 2018; Kong, 2021) and other techniques (Avila, 2018; Ramos, 2016) or using the IVs determined by the best RA model as the IVs used in MLP and SVR. Additionally, other ML methods could be tested for prediction efficacy.

Additionally, the use of more complex RA models, such as models with loops and state-based models, could be investigated. Prior research on electricity data has shown that different ML methods have certain advantages and disadvantages given the particular application (Sheraz, 2021). Further extensions could also refine the Sunrise/Sunset variable to capture the optimal Sunrise/Sunset timeframe for each WEIM participant where solar penetration is the highest, or more acutely a focus on the window of time of peak net load ramp or decline.

Afterword

Summary of Original Contributions

Paper 1 develops the joint lattice of RA and BN general graphs for four variables which increases the number of general graphs with unique independence structures from 20 in the four variable RA lattice and 20 in the four variable BN lattice to 30 in the joint RA-BN lattice, and when variable labels are added, increases the number of unique specific graphs from 114 in the RA lattice and 185 in the BN lattice to 238 in the joint lattice.

The integration of the two lattices offers a richer and more expansive way to model complex systems leveraging the V-structure unique to BN graphs and the allowability of cycles in RA graphs. The joint RA-BN lattice of general graphs presented in this paper expands the set of general graphs with unique independence structures (or, equivalently, with unique interdependence structures) beyond what was previously available by either RA alone or BN alone, thus allowing for representations of complex systems which are (i) more accurate relative to data and/or (ii) simpler and thus more comprehensible and more generalizable than would be possible by modeling only with RA or only with BN. This joint lattice thus demonstrates how these two related frameworks – RA and BN – both members of the family of probabilistic graphical modeling methodologies, can be integrated into a unified framework.

Paper 2 builds on Paper 1, which developed the BN neutral system *general* graph lattice, expanding it to offer the BN neutral system *specific* graph lattice. Paper 2 also

builds on the joint RA-BN neutral system general graph lattice developed from Paper 1, expanding it to offer the *joint* RA-BN neutral system *specific* graph lattice. In developing these new lattices, Paper 2 extends RA notation to encompass BN graphs.

For four variables, the joint RA-BN neutral system specific graph lattice developed in Paper 2 increases the number of unique specific graphs from 114 in the RA lattice and 185 in the BN lattice to 238 in the joint lattice. The integration of the two lattices offers a richer and more expansive way to model and represent complex systems leveraging the V-structure unique to BN graphs and the ability to accommodate loops and hypergraphs in the RA lattice.

Paper 2 also develops an algorithm to generate the joint RA-BN neutral system general and specific graph lattices for any number of variables in both upward and downward directions. The exhaustive and non-redundant RA and BN lattices follow the more general Rho lattice proposed by Klir (1985).

Even though Papers 1 and 2 focus on graphs with only four variables, the findings are general, and apply to any number of variables. This general applicability applies to both the lattice integration and the search algorithm that was developed. The four variable analyses are sufficient to establish the formal relationship between RA and BN. Performing an exhaustive analysis for five or more variables is not only unnecessary but would also be too burdensome because of the exponential combinatorial expansion that would result.

Paper 2 also develops the RA augmented *directed* system lattice for four variables, which is a new extension of the conventional RA directed system lattice. While the conventional RA directed system lattice encompasses all prediction graphs in the BN directed system lattice (under the restriction that DVs in BN models are not parent variables in V-structures), the RA conventional directed system lattice did not include naïve Bayes graphs. Doing so increases the number of general graphs from nine in the conventional RA lattice to 12 in the augmented lattice, and the number of specific graphs from 19 to 31. The augmented RA directed system lattice thus offers more candidate graphs, and this allows for the possibility of more accurate or simpler and thus more generalizable RA prediction models.

Paper 2 also develops the BN directed system lattice of prediction graphs for four variables, reducing the number of possible specific graphs from 185 in the BN neutral system lattice to 18 in the BN directed system lattice—a significant compression of the BN neutral system lattice when prediction of a single DV is the goal. Paper 2 also shows that all of the graphs in the BN directed system lattice (where this lattice disallows graphs where the DV is a parent in a V-structure) are equivalent in their predictions to RA graphs, although many of them have fewer degrees freedom than their RA-equivalent counterpart. The augmented RA directed system lattice thus encompasses all of the BN directed system general graphs in terms of prediction, and offers additional predicative graphs, those including loops, that are not in the BN lattice.

Paper 3 has as one of its primary aims the application of machine learning methods to build a capacity prediction model that improves upon the current Industry Model. The results show that the best RA Model does in fact do that. The best RA Model reduces the required total reserves relative to the Industry Model by 150.9 MW on average, a reduction of 25.4%. INC reserves are reduced on average 62.7 MW, a reduction of 23.0% relative to the Industry Model, while DEC reserves are reduced on average 88.2 MW, a reduction of 27.3. These results are for one of nineteen WEIM participants. If the RA Model was used for all participants, it would result in a significant reduction in capacity reserves that would need to be held. For BPA alone, a 62.7 MW reduction of INC reserves is enough to power approximately 25,000-50,000 homes. Conservatively, reducing the reserve requirement by this amount would result in approximately \$4 million¹² in annual savings for the Bonneville Power Administration which is 1 of 19 WEIM participants and which was the focus of the data analysis in Paper 3.

Paper 3 also reports a comparison of point estimate prediction results for four different machine learning methods. Among these methods, RA performed significantly better than the other ML methods. Based on the R squared metric, RA resulted in 33.8%, MLP 18.1%, BN 13.9%, SVR-rbf 7.5% and the Industry Model 7.5%. Our presumption is that RA performed better than the other methods because it is able to

¹² INC savings would be $\$168/\text{MW-day} * 365 \text{ days} * 62.7 \text{ MW} = \3.84 million . DEC savings would be $\$12/\text{MW-day} * 365 \text{ days} * 88.2 \text{ MW} = \386 thousand .

efficiently determine the optimal set of IVs to be used in the model whereas MLP and SVR used f-regression to select the predictive IVs. We do not rule out the possibility that the performance of MLP and SVR might have been improved if a better feature selection procedure had been employed. It should be noted, however, that feature selection is built in to RA: it's the result of using the simplest type of RA models, namely variable-based models without loops. The BN algorithm gave results inferior to RA in part because its algorithms are not optimal for exploratory search and in part because network complexity is limited by the computational cost of the EM algorithm which is used to make predictions about missing values or missing states.

Paper 3 also discusses the results and theoretical comparisons between RA and BN as they are both machine learning methods that are analytically very similar as shown extensively in Papers 1 and 2. Paper 3 built on prior research from Paper's 1 and 2 and used a concrete example to show that the best RA Model identified in Paper 3 can be replicated exactly in a BN. However the BN search algorithm, in particular the computational cost of the EM algorithm, prevent such complex BN models from being found by standard BN search algorithms. Further, Paper 3 used toy examples to show that several BNs with different edge topologies are analytically equivalent to a single RA graph, these toy examples further amplify and illustrate the prior work from Paper 2.

In Paper 3, the best RA Model resulted in 15 of the 22 possible IVs being included in the model. Of these, the most predictive variables are hour of day (HourOfDayx2hrxgroups), 48 hour wind forecast (VERFCFHFrtegt), and Sunrise/Sunset

(SunriseSunset). These variables were also found and included in the best SVM and MLP models. The 48 hour wind forecast and Sunrise/Sunset variable reflect times of energy imbalance uncertainty, thus their information content is useful in making predictions about the INC and DEC reserves requirement. Also interesting, the best RA Model, and each best model from all 5 folds included the 48 hour wind forecast, 96 hour wind forecast and 120 hour wind forecast. This is an indication that there is additional, significant information, contained in the 96 and 120 hour wind forecasts not contained in the 48 hour wind forecast.

Summary of Possible Research Extensions

Future extensions of the research identified in Paper 1 were carried out in Paper 2. Specifically, the paper pointed to the desirability of developing an algorithm to search the joint RA-BN lattice. Paper 1 was also restricted to neutral systems, and pointed to the desirability of analyzing RA and BN directed system prediction graphs. These two tasks that were not done in paper 1 were later accomplished in paper 2.

Extensions identified in Paper 2 centered on advancing the lattice search algorithm to generate a hierarchical nesting of graphs to allow statistical significance testing at each incremental lattice level as well as making the algorithm more efficient in its search of optimal models. The algorithm developed in Paper 2 results in an exhaustive search, which is much more computationally demanding than a more efficient search algorithm that may optimize based upon a pre-determined scoring metric. One possible search approach could mimic the RA beam search, where the best

scoring models at the next higher or lower level (depending on whether the search is upward or downward through the lattice) would be trimmed based on a defined scoring metric. Only the highest scoring models would be kept. This would significantly reduce the computation cost and time of searching the lattice for the best representative model. Further, an interesting extension could be to generate a BN lattice of partially directed acyclic graphs and compare that to the RA lattice.

Another promising extension of the theoretical work in Paper 2 would be to develop hybrid RA-BN structures. It is possible to create such hybrid RA-BN graphs that have mathematical structures that are different than those found in RA or BN, and these would add to the current joint RA-BN lattice, offering an even greater expansion of possible complex system representative structures. Such structures would extend BNs to allow for hypergraphs or cycles which are both common to RA and extend RA to allow for V-structures, whose independence relations are not encompassed in RA models. A further extension could look at state-based RA and investigate how it relates to structures common to both RA and BN and also to structures that are unique to BN.

Paper 2 also showed that BN general graph representations use a single representative member from the class of equivalent general graphs. An extension could be to develop a representation of the lattice that instead uses a single Partially Directed Acyclic Graph that has some non-directed edges to represent the whole equivalence class.

Further, Paper 2 showed that certain BN graphs are equivalent in terms of prediction to certain RA graphs, but that the BN graph can do so with less degrees of freedom. An extension of this work could analyze whether this reduction in degrees of freedom offers an advantage in calculations of statistical significance.

A related issue that could benefit from additional research is to integrate statistical tests of augmented directed RA models whose reference posits independence between IVs with statistical tests of standard directed RA models whose (conventional) reference posits all the IVs joined in a single relation. This is necessary because augmented models whose reference posits independence between IVs are not hierarchically nested with conventional directed system models whose IVs are joined in a single relation. If a directed system (augmented) model were statistically significant relative to a reference that posits inter-IV independence, it might be possible to do additional tests to discern if this difference was due to the presence in the model of dependence between the IVs or to the presence in the model of statistically significant predictive relations between the IVs and the DV. A further extension would be to develop an efficient search algorithm to generate and search the augmented RA directed system lattice.

Other possible extensions in Paper 2 include relaxing the restriction that does not allow DVs to be parent nodes in V-structures. This extension would expand the BN directed system lattice to include such unusual BN predictive graphs. As shown in Figure 39, relaxing this restriction results in models that have unique probability distributions,

and thus make unique predictions not found in the conventional BN directed system lattice. Thus, this extension would expand the lattice of candidate graphs.

Pearl (2000) argues that BN models can represent causal interactions, not merely statistical associations. This could be investigated in the context of RA, to see under what conditions similar claims might be made for some RA models. Studený shows that there is no system that completely describes all conditional independence statements (Studený, 1992, 2005) however his method of structural imsets has known computational limitations when variables exceed five (Bouckaert and Studený, 2005). The new contributions in this dissertation that combine RA and BN lattices to more broadly express conditional independencies may benefit from further exploring Studený's work.

Paper 3 applied machine learning methods to predict electricity capacity requirements. It wasn't the purpose of this study to do a systematic comparison of methods, but doing such a study would be a desirable extension of the limited study done in this dissertation. From a methods perspective, future extensions or improvements would assess additional pre-processing techniques for MLP and SVR such as principal component analysis to ensure the optimal set of IVs is used in these models as well as more advanced implementations of MLP, for example testing the optimal number of hidden layers. Another related extension would be to give MLP and SVR the IVs that RA found as a form of preprocessing, and see how well MLP and SVR do with the larger set of IVs that RA actually used. Other machine learning methods could be

also tested, such as decision trees, genetic algorithms, and regression models and one could also investigate how continuous variables are handled in the BN literature and try to apply this to RA. Testing these other methods was not the focus of the dissertation (the focus was primarily RA and BN), but more extensive comparison of RA/BN to other methods could be done as well as to the continuous application of BNs.

Further, for RA, only loopless models were used, whereas past RA work (Froemke, 2017, Nunes, 2020) generally found that models with loops performed better on training data. However, this study differs from past work in dividing data into three parts – training, validation, and test – and selecting models based on their performance in the validation data. Loopless models performed well because the sample size was large, resulting in loopless models that generalized better to the validation data than models with loops picked by metrics such as BIC or AIC. This could be explored to understand this finding better. Additionally in this study a backup model was used to make RA predictions for missing IV states, and specific requirements were adopted for the choice of this backup model. Other sets of requirements could be explored to investigate what might be the optimum backup model for predicting missing IV states. (The RA OCCAM software allows the user to identify a backup model, but currently provides no assistance in its choice.) Conceivably BN software might be modified to allow users to employ the fast RA backup model procedure instead of the much slower EM algorithm.

Further, additional extensions of Paper 3 could include adding or refining existing variables to more accurately represent the appropriate signal. For example, improving the Sunrise/Sunset variable to capture the optimal time of highest solar generation ramping and decline, or more acutely a focus on the window of time of peak net load ramp or decline. The predictive modeling that was done in Paper 3 could be adopted by the industry, and as shown, would result in a significant (25.4%) reduction in the INC and DEC capacity requirement for BPA. Adoption would result in material energy cost savings to end use consumers. There are some barriers to adopting machine learning models for this type of application. The first is an understanding by decision makers of the models results and limitations. RA and BN both identify the variables that are being used to make predictions and for each IV state the probability of the DV, given the IVs is known, and can be investigated for understanding. This is not common among all machine learning methods, like MLP or SVR, where variables are known, but underlying weightings are difficult to interpret. Another challenge generally for predictions on the electrical grid, and for any critical infrastructure operation, is the fact that if predictions are wrong, they can have severe consequences. The impact of error is much more significant than in other industries where human safety and wide spread negative economic impact are not at play. Thus to fully implement machine learning methods for this type of application, further education of decision makers is likely needed on how these models make predictions, how well they generalize to new data, and the potential consequences of these models being in error. Further, if adopted, it would not be

unreasonable to start with a conservative, less complex model, or to start by running a model in parallel to the conventional approach to further validate the model results and gain comfort with its accuracy and generalizability.

References

1. Andersson, S.; Madigan, D.; Perlman, M. A Characterization of Markov Equivalence Classes For Acyclic Digraphs. *Ann. Stat.* 1997, *25*, 505–541.
2. Ashby, W.R. Constraint Analysis of Many-Dimensional Relations. *Gen. Syst. Yearb.* 1964, *9*, 99–105.
3. Avila, N.F.; Figueroa, G.; Chu, C.-C. Ntl detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting. *IEEE Trans. Power Syst.* 2018, *33*, 7171–7180.
4. Bouckaert, R. Properties of Bayesian Belief Network Learning Algorithms. In Proceedings of the Tenth international conference on Uncertainty in artificial intelligence, *Seattle, WA, U.S.A., July 29-31, 1994*; pp. 102–109.
5. Bouckaert, R.R.; Studený, M. Racing for Conditional Independence Inference. In: Godo, L. (eds) Symbolic and Quantitative Approaches to Reasoning with Uncertainty. ECSQARU 2005. Lecture Notes in Computer Science(), vol 3571. Springer, Berlin, Heidelberg. 2005. https://doi.org/10.1007/11518655_20
6. Broekstra, G. Nonprobabilistic constraint analysis and a two stage approximation method of structure identification. In Proceedings of the 23rd Annual SGSR Meeting, Houston, TX, January 3-8, 1979.
7. Buntine, W. Classifiers: A theoretical and empirical study. In Proceedings of the IJCAI, Sydney, Australia, August 24-30, 1991; Morgan Kaufmann: *Burlington, MA, USA*, 1991; pp. 638–644.
8. Buntine, W. Theory refinement on Bayesian networks. In Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence, Los Angeles, CA, USA, July 13-15, 1991, pp. 52–60.
9. Buzau, M.M.; Tejedor-Aguilera, J.; Cruz-Romero, P.; Gomez-Exposito, A. Detection of non-technical losses using smart meter data and supervised learning. *IEEE Trans. Smart Grid* 2018, *10*, 2661–2670.
10. Cavallo, R. *The Role of System Science Methodology in Social Science Research*; Martinus Nijhoff Publishing: *Leiden*, The Netherlands, 1979.
11. Chen, Y. Structure Discovery in Bayesian Networks: Algorithms and Applications. Master Thesis. Iowa State University, Ames, IA, USA, 2016.
12. Chickering, D. Transformational Characterization of Equivalent Bayesian Network Structures. In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montréal Québec Canada, August 18-20, 1995; pp.87–98.
13. Chickering, D. Learning Equivalence Classes of Bayesian-Network Structures. *J. Mach. Learn. Res.* 2002, *2*, 445–498.
14. Chickering, D.; Geiger, D.; Heckerman, D. Learning Bayesian networks: Search methods and experimental results. In Proceedings of the Fifth Conference on

- Artificial Intelligence and Statistics, Society for Artificial Intelligence and Statistics: Ft. Lauderdale, FL, USA, January 4-7,1995; pp. 112–128.
15. Chickering, D.; Heckerman, D.; Meek, C. A Bayesian approach to learning Bayesian networks with local structure. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*; Geiger, D., Ed.; Morgan Kaufmann: Providence, RI, USA, August 1, 1997; pp. 80–90.
 16. Chickering, D.; Heckerman, D.; Meek, C. Large-Sample Learning of Bayesian Networks is NP-Hard. *J. Mach. Learn. Res.* 2004, 5, 1287–1330.
 17. Conant, R. Mechanisms of intelligence: Ashby's writings on cybernetics. In *Intersystems Publications, Seaside, CA, U.S.A*; 1981.
 18. Conant, R. Extended dependency analysis of large systems. *Int. J. Gen. Syst.* 1988, 14, 97–123.
 19. Cooper, D.A. A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* 1992, 9, 309–347.
 20. Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B.* 39 (1): 1–38. JSTOR 2984875. MR 0501537.
 21. Driver, E.; Morrell, D. Implementation of Continuous Bayesian Networks Using Sums of Weighted Gaussians. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montréal, Québec, Canada, August 18 – 20, 1995.*
 22. Drucker, Harris; Burges, Christ. C.; Kaufman, Linda; Smola, Alexander J.; and Vapnik, Vladimir N. (1997); "Support Vector Regression Machines", in *Advances in Neural Information Processing Systems 9, NIPS 1996, 155–161, MIT Press.*
 23. Forre, P.; Mooij, J. Causal Calculus in the Presence of Cycles, Latent Confounders and Selection Bias. In *Proceedings of Uncertainty in Artificial Intelligence Conference, Tel Aviv, Israel, July 22-25, 2019 ; Volume 115, pp. 71–80.*
 24. Friedman, N.; Goldszmidt, M. Building classifiers using Bayesian Networks. In *Proceedings of the national conference on artificial intelligence, Portland, OR, U.S.A, August 4-8, 1996.*
 25. Friedman, N.; Koller, D. Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. *Mach. Learn.* 2003, 50, 95–125.
 26. Froemke, Cecily Corrine and Zwick, Martin, "Predicting Risk of Adverse Outcomes in Knee Replacement Surgery with Reconstructability Analysis" (2017). *Systems Science Faculty Publications and Presentations.* 127.
https://pdxscholar.library.pdx.edu/sysc_fac/127.
 27. *Genie Modeler User Manual. Version 3.0.R2. 2020. BayesFusion, LLC.*
(<https://support.bayesfusion.com/docs/GeNle.pdf>).

28. Ghosal, M.; Campbell, A.M.; Elizondo, M.A.; Samaan, N.A.; Nguyen, Q.H.; Nguyen, T.B.; Muñoz, C.; Hernández, D.M. Grid Reserve and Flexibility Planning Tool (GRAF-Plan) for Assessing Resource Balancing Capability under High Renewable Penetration. *IEEE Open Access J. Power Energy* 2022. <https://doi.org/10.1109/OAJPE.2022.3169729>.
29. Gillispie, S.; Perlman, M. Enumerating Markov Equivalence Classes of Acyclic Digraph Models. In *Proceedings of the Seventeenth conference on Uncertainty in Artificial Intelligence*, Seattle, WA, U.S.A, August 2, 2001; pp. 171–177.
30. Harris, M.; Zwick, M. Joint Lattice of Reconstructability Analysis and Bayesian Network General Graphs. In *Proceedings of the Tenth International Conference on Complex Systems*; Springer: Cham, Switzerland, 2021.
31. Harris, M.; Zwick, M. Graphical Models in Reconstructability Analysis and Bayesian Networks. *Entropy* 2021, 23, 986. <https://doi.org/10.3390/e23080986>.
32. Harris M, Kirby E, Agrawal A, Pokharel R, Puyleart F, Zwick M. Machine Learning Predictions of Electricity Capacity. *Energies*. 2023; 16(1):187. <https://doi.org/10.3390/en16010187Harris>.
33. Heckerman, D.; Geiger, D.; Chickering, D. Learning Bayesian networks: The combination of knowledge and statistical data. In *Proceedings of The Tenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann: Seattle, WA, USA, July 29-31, 1994; pp. 293–301.
34. Javidian, M.A.; Wang, Z.; Lu, L.; Valtorta, M. On a hypergraph probabilistic graphical model. *Ann. Math Artif. Intell.* 2020, 88, 1003–1033, doi:10.1007/s10472–020-09701–7
35. Jost, E.; Braun, A.; Fritz, R. Dynamic dimensioning of frequency restoration reserve capacity based on quantile regression. In *Proceedings of the 2015 12th International Conference on the European Energy Market (EEM)*, Lisbon, Portugal, 19–22 May 2015; pp. 1–5. <https://doi.org/10.1109/EEM.2015.7216769>.
36. Klir, G. Identification of generative structures in empirical data. *Int. J. Gen. Syst.* 1976, 3, 89–104.
37. Klir, G. *The Architecture of Systems Problem Solving*; Plenum Press: New York, NY, USA, 1985.
38. Klir, G. Reconstructability analysis: An offspring of Ashby's constraint theory. *Syst. Res.* 1986, 3 267–271.
39. Koivisto, M.; Sood, K. Exact Bayesian structure discovery in Bayesian networks. *J. Mach. Learn. Res.* 2004, 5, 549–573.
40. Kong, X.; Zhao, X.; Liu, C.; Li, Q.; Dong, D.; Li, Y. Electricity theft detection in lowvoltage stations based on similarity measure and dt-ksvm. *Int. J. Electr. Power Energy Syst.* 2021, 125, 106544.

41. Krippendorff, K. On the identification of structures in multivariate data by the spectral analysis of relations. In *General Systems Research: A Science, a Methodology, a Technology*; Gaines, B.R., Ed.; Society for General Systems Research: Louisville, KY, USA, 1979; pp. 82–91. Available online: http://repository.upenn.edu/asc_papers/207.
42. Krippendorff, K. An algorithm for identifying structural models of multivariate data. *Int. J. Gen. Syst.* 1981, 7, 63–79.
43. Krippendorff, K. *Information Theory: Structural Models for Qualitative Data; Quantitative Applications in the Social Sciences #62*; Sage: Beverly Hills, CA, USA, 1986.
44. Larranaga, P.; Kuijpers, C.; Murga, R.; Yurramendi, Y. Learning Bayesian network structures by searching for the best ordering with genetic algorithms. *IEEE Trans. Syst. Man Cybern.* 1996, 26, 487–493.
45. Lauritzen, S. *Graphical Models*; Oxford Statistical Science Series; Oxford University Press: New York, NY, USA, 1996.
46. Lu, S.; Makarov, Y.V.; Brothers, A.J.; McKinstry, C.A.; Jin, S.; Pease, J.H. Prediction of power system balancing requirement and tail event. IEEE PES TD 2010, 2010, 1–7. <https://doi.org/10.1109/TDC.2010.5484347>.
47. Malone, B.; Yuan, C.; Hanse, E. Memory-Efficient Dynamic Programming for Learning Optimal Bayesian Networks. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA, U.S.A., August 7–11, 2011; pp. 1057–1062.
48. Meek, C. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*; Hanks, S., Besnard, P., Eds.; Morgan Kaufmann: Burlington, MA, USA, 1995; pp. 403–410.
49. MIT. D-Separation. 2015. Available online: <http://web.mit.edu/jmn/www/6.034/d-separation.pdf>.
50. Murphy, K. A Brief Introduction to Graphical Models and Bayesian Networks. 1998. Available online: https://www.cs.ubc.ca/~murphyk/Bayes/bayes_tutorial.pdf.
51. Neapolitan, R. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*; Wiley: New York, NY, USA, 1989.
52. Nunes, A.; Zwick, M.; Wakeland, W. Sensitivity analysis of an agent-based simulation model using reconstructability analysis. *International Journal of General Systems*. 2020. <https://doi.org/10.1080/03081079.2021.1874947>.
53. Nuclear Regulatory Commission (NRC). What is a Megawatt? <https://www.nrc.gov/docs/ML1209/ML120960701.pdf>. 2012.

54. Obando-Ceron, J.S.; Moreno-Chuquen, R. Quantification of Operating Reserves with Wind Power in Day-Ahead Dispatching. 2018 IEEE ANDESCON, Santiago de Cali, Colombia, 22–24 August 2018; pp. 1–5.
<https://doi.org/10.1109/ANDESCON.2018.8564589>.
55. Pearl, J. Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In Proceedings of the 7th Conference of the Cognitive Science Society, Irvine, CA, USA, August 15–17, 1985; pp. 329–334.
56. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*; Morgan Kaufmann Publishers, Inc: San Francisco, CA, 1988.
57. Pearl, J. *Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2000.
58. Pearl, J, Verma T. The logic of representing dependencies by directed graphs. In Proceedings of the 6th National Conference on Artificial Intelligence, Seattle, WA, USA, July 13–17, 1987.
59. Pedregosa et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12, October 2011: pp. 2825–2830.
60. Ramos, C.C.; Rodrigues, D.; de Souza, A.; Papa, J. On the study of commercial losses in brazil: A binary black hole algorithm for theft characterization. IEEE Trans. Smart Grid 2016, 9, 676–683.
61. Rebane, G.; Pearl, J. The recovery of causal polytrees from statistical data. In Proceedings of the Third Conference on Uncertainty Artificial Intelligence, July 10, 1987; pp. 222–228.
62. Richardson, T, Spirtes, P. Ancestral graph Markov models. *Ann. Stat.* 2002, 30, 962–1030
63. Rubin, D. Bayesian Inference for Causal Effects: The Role of Randomization. *Ann. Stat.* 1978, 6, 34–58.
64. Sheraz, A.; Herodotos, H.; Mohsin, M.S.; Javaid, N.; Ashraf, N.; Aslam, S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sustain. Energy Rev.* 2021, 144, 110992. ISSN 13640321. <https://doi.org/10.1016/j.rser.2021.110992>.
65. Studený, M. Conditional independence relations have no finite complete characterization. In: Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, volume B, Academia, Prague 1992, pp. 377–396.
66. Studený, M. Probabilistic Conditional Independence Structures. Springer-Verlag, London, 2005.
67. Studený, M.; Vomlel, J. Geometric view on learning Bayesian Network Structures. *Int. J. Approx. Reason.* 2010, 51, 573–586.

68. Tang , Y.; Srihari, N. Efficient and Accurate Learning of Bayesian Networks using Chi-Squared. In Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, November 11-15, 2012.
69. Tian, J.; He, R.; Ram, L. Bayesian model averaging using the k-best Bayesian network structures. In Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, U.S.A., July 8-11, 2010.
70. Vapnik, V., 1995. *The nature of statistical learning theory*. Springer science & business media.
71. Verma T, Pearl J. Equivalence and synthesis of causal models. In Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, U.S.A., July 27-29, 1990; pp. 255–270.
72. Werbos, P., 1974. Beyond regression:" new tools for prediction and analysis in the behavioral sciences. Ph. D. dissertation, Harvard University.
73. Wright, S. Correlation and causation. *J. Agric. Res.* 1921, Vol. 20, 557–585.
74. Wright, S. The method of path coefficients. *Ann. Math. Stat.* 1934, 5, 161–215.
75. Willet, K.; Zwick, M. A software architecture for reconstructability analysis. *Kybernetes* 2004, 33, 997–1008. Available online: https://works.bepress.com/martin_zwick/55/
76. Zhang, H. The Optimality of Naive Bayes. In Proceedings of the FLAIRS conference, Miami Beach, FL, U.S.A., May 12-14, 2004.
77. Zwick, M. Reconstructability Analysis of Epistasis. *Ann. Hum. Genet.* 2010, 157–171. doi:10.1111/j.1469-1809.2010.00628.x. Available online: https://works.bepress.com/martin_zwick/3/.
78. Zwick, M. Wholes and parts in general systems methodology. In *The Character Concept in Evolutionary Biology*; Wagner, G. Ed.; Academic Press: New York, NY, USA, 2001. Available online: https://works.bepress.com/martin_zwick/52/.
79. Zwick, M. An overview of reconstructability analysis. *Kybernetes* 2004, 33, 887–905. Available online: https://works.bepress.com/martin_zwick/57/.
80. Zwick, M.; Johnson, M.S. State-based reconstructability analysis. *Kybernetes*, 2004, 33, 1041–1052. Available online: https://works.bepress.com/martin_zwick/47/.
81. Zwick, M. Carney, N.; Nettleton, R. Exploratory Reconstructability Analysis of Accident TBI Data. *Int. J. Gen. Syst.* 2018, 47, 174–191. doi:10.1080/03081079.2017.1412435. Available online: https://works.bepress.com/martin_zwick/80/.
82. Zwick, M., Fusion, J., Wilmot, B. Reconstructability Analysis of Epistatic Functions. *Journal of Molecular Engineering & Systems Biology* 2012, <http://www.hoajonline.com/journals/pdf/2050-1412-1-4.pdf>

Appendices

Appendix A.1 RA Loop Detection Procedure

The contents of Appendix A.1 and A.2 are the work of other researchers (Krippendorff, 1986; MIT, 2015) and are included here to allow Paper 2 to be understood in a self-contained way.

In the RA graphs of Figure 6 and Figure 10 of Paper 2, graphs that do not have loops are highlighted in bold, while graphs that have loops are non-bold. Graphs without loops are fitted algebraically, whereas graphs with loops are fitted using iterative proportional fitting. To determine if a graph has a loop, the following procedure is performed:

Given the set of relations of a specific graph:

1. Remove all variables that are unique to any individual relation
2. Remove any relation that is equal to or embedded in any other relation of the (remaining) set
3. Repeat 1 and 2 until either
4. No variables remain, in which case there are no loops, or
5. The remainder is unalterable by steps 1 or 2, in which case there are loops

For example, in Figure 10, graph G7, illustrated by specific graph ABC:ABD does not have a loop. Krippendorff's loop detection algorithm (Krippendorff, 1986) produces the following results. First, removing variables unique to both ABC and ABD removes C from ABC and D from ABD. What remains is AB:AB, for which the second AB is redundant and

thus removed, leaving AB. Then, removing the variables unique to AB removes both A and B, leaving the null set, and thus this specific graph does not contain a loop.

By contrast, in Figure 10, graph G8, illustrated by specific graph ABC:AD:BD, does have a loop. Krippendorff's loop detection algorithm (Krippendorff, 1986) produces the following results. First, removing variables unique to one relation removes C from ABC. What remains is AB:AD:BD. There is no redundant relation, i.e., no relation that is repeated or embedded in another relation. There are also no variables unique to one relation. Therefore, nothing further can be removed; because the remaining set is unalterable, the specific graph has a loop.

Appendix A.2 D-Separation Procedure

The following provides the procedure for determining all independencies for a BN (MIT, 2015).

Step 1. List all possible independence statements for a given BN.

Step 2. For each independence statement, construct the 'ancestral graph' (Rebane, 1987) for the variables mentioned in the independence statement.

Step 3. 'Moralize' the ancestral graph by adding an undirected edge between two nodes if they have a common child.

Step 4. 'Disorient' the moralized, ancestral graph, by making all edges undirected.

Step 5. Delete the givens (nodes) and any of their edges from the independence statement being tested.

Step 6. Read the answer to the independence statement question from the remaining graph, if the variables are disconnected in the remaining graph, the answer to the independence statement is in the affirmative.

The following provides examples of the D-separation procedure for BN12 and BN9*:

Example 1

Step 1. List all possible independence statements for a given BN. For four variables, Table 11 is the complete list.

Step 2. For each independence statement, construct the ‘ancestral graph’ (Rebane, 1987) for the variables mentioned in the independence statement.

An ancestral graph of the probability expression includes all nodes listed in the independence statement that is being tested and all parents, grandparents, great-grandparents, etc., of those nodes.

BN12 and the independence statement $(A \perp B \mid D)$ will be used as an example throughout the remaining procedure (Steps 2–5, Figures 49-52).

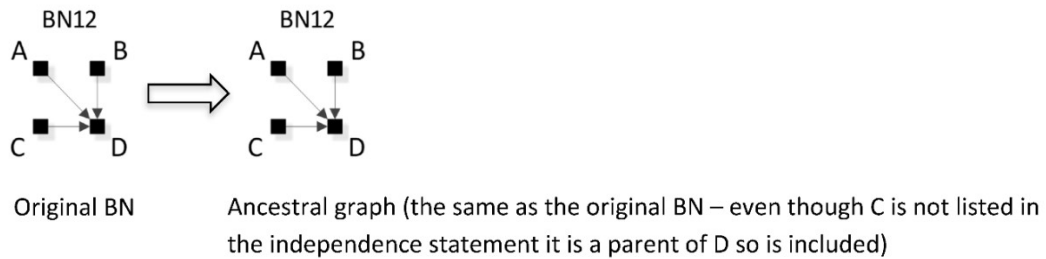


Figure 49 Step 2, create the ancestral graph

Step 3. ‘Moralize’ the ancestral graph by adding an undirected edge between two nodes if they have a common child.

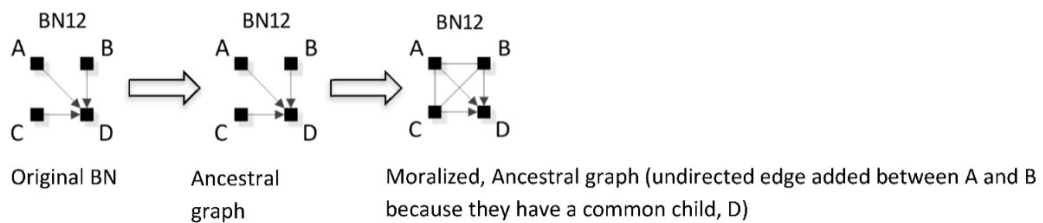


Figure 50 Step 3, moralize the ancestral graph

Step 4. 'Disorient' the moralized ancestral graph by making all edges undirected.

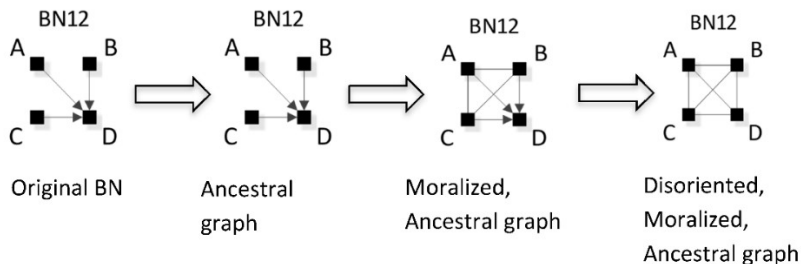


Figure 51 Step 4, disorient the moralized, ancestral graph

Step 5. Delete the givens from and any of their edges. In the continuing example,

D is the 'given' ($A \perp B \mid D$), thus D and its connected edges to A, B, and C are removed.

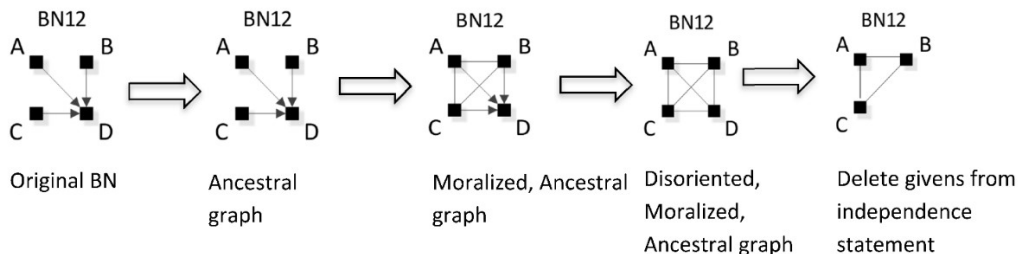


Figure 52 Step 5, delete the givens

Step 6. Read the answer to the independence statement question from the remaining graph, if the variables are disconnected in the remaining graph, the answer to the independence statement is in the affirmative. In this example, the independence statement being tested is the assertion that A and B are independent given D. This assertion is false because A and B are connected in the remaining graph; thus, they are not *conditionally* independent given D.

Example 2

Consider a second example, using BN9*. The independence statement being tested here is the assertion: C independent of D given A and B ($C \perp D | A, B$). Figure 53 shows all steps in the procedure for this example, affirming C and D are indeed independent given A and B.

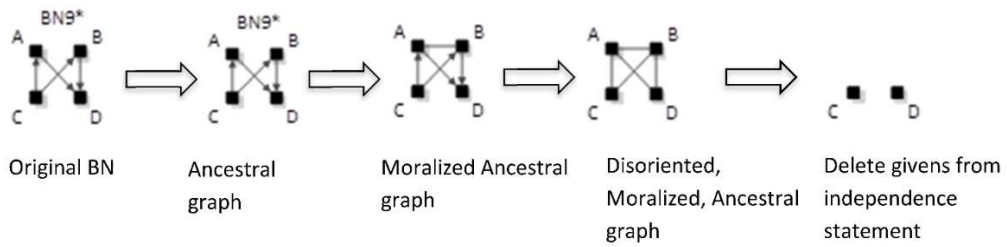


Figure 53 Example, full D-separation procedure for independence statement ($C \perp D | A, B$)

Appendix B Input parameters for RA, BN, SVR, and MLP analysis in Chapter 3

Reconstructability Analysis

| | |
|---------------|-------|
| Program | OCCAM |
| Search Width | 3 |
| Search Levels | 50 |
| Loops | No |

Bayesian Networks

| | |
|------------------------|----------------------------|
| Program | Genie |
| BN Type | Augmented Bayesian Network |
| Link Probability | 0.1 |
| Prior Link Probability | 0.001 |
| Max Parent Count | 22 |

Support Vector Regression

| | |
|-----------------|--|
| Program | Python Sklearn |
| Preprocessing | MinMax Scalar |
| Hyper-parameter | default values |
| Kernels | radial basis function, linear, polynomial, and sigmoid |

Multi-layer Perceptron

| | |
|-----------------|-------------------------|
| Program | Python Sklearn |
| Preprocessing | MinMax Scalar |
| Hyper-parameter | default values |
| Hidden Layers | 2 with 100 neurons each |