7-2018

# Preliminary Results of Bayesian Networks and Reconstructability Analysis Applied to the Electric Grid

Marcus Harris
*Portland State University*

Martin Zwick
*Portland State University*, zwick@pdx.edu

# Preliminary results of Bayesian Networks and Reconstructability Analysis applied to the electric grid

*Portland State University, Systems Science*

*Marcus Harris*

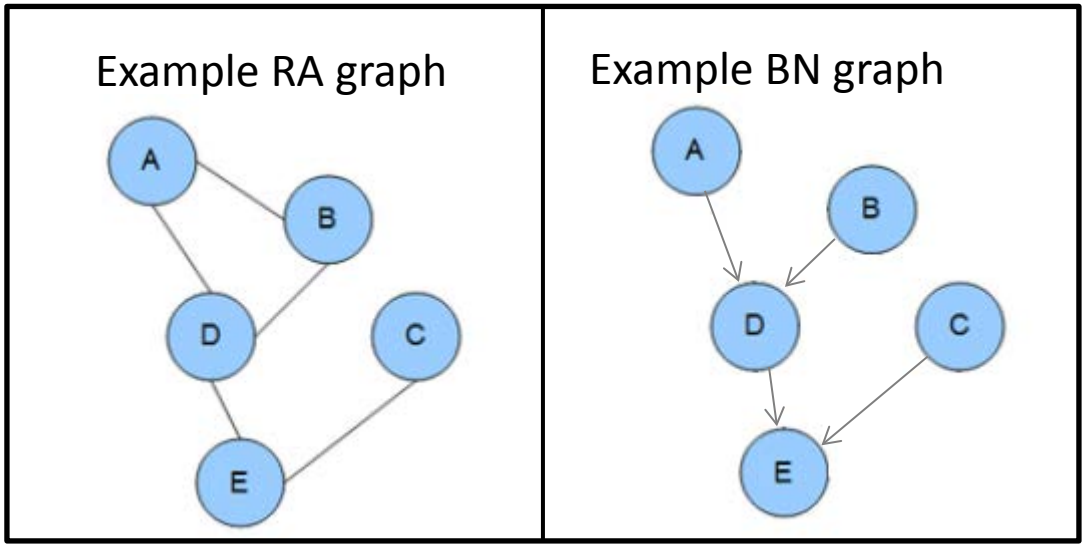
7/26/2018

# Outline

1. Brief background on RA & BN models
2. Brief overview of the application of RA & BN to build prediction models for comparison
3. Preliminary modeling results and possible future research

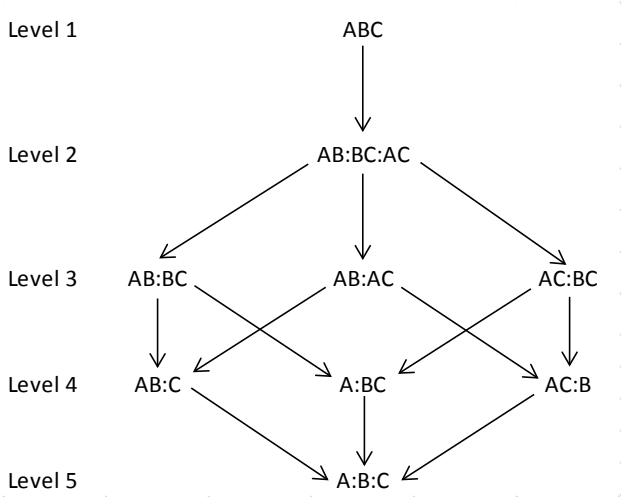
# RA & BN background

- RA is an analytical approach developed in the systems community (Ashby 1964) that combines graph theory and information theory.
  - Graph theory provides the structure of relations (model of the data) between variables and information theory characterizes the strength and the nature of the relations.
- BN are another graphical modeling approach for data analysis and knowledge discovery established as a field of study in the late 1980's separately by Pearl and Neapolitan.
  - As does RA, BN combines graph and probability theory; in both, graph theory provides the structure and probability theory characterizes the nature of relationships between variables.
- RA graph structures are undirected and BN graph structures are directed. RA graph structures allow loops whereas BN does not; most structures have loops so this is a limitation of BN. BN models allow the hypothesis of IV independence, which RA models do not. Also BN implementations explicitly address the problem of missing data
- Both BN and RA models can be used for prediction.



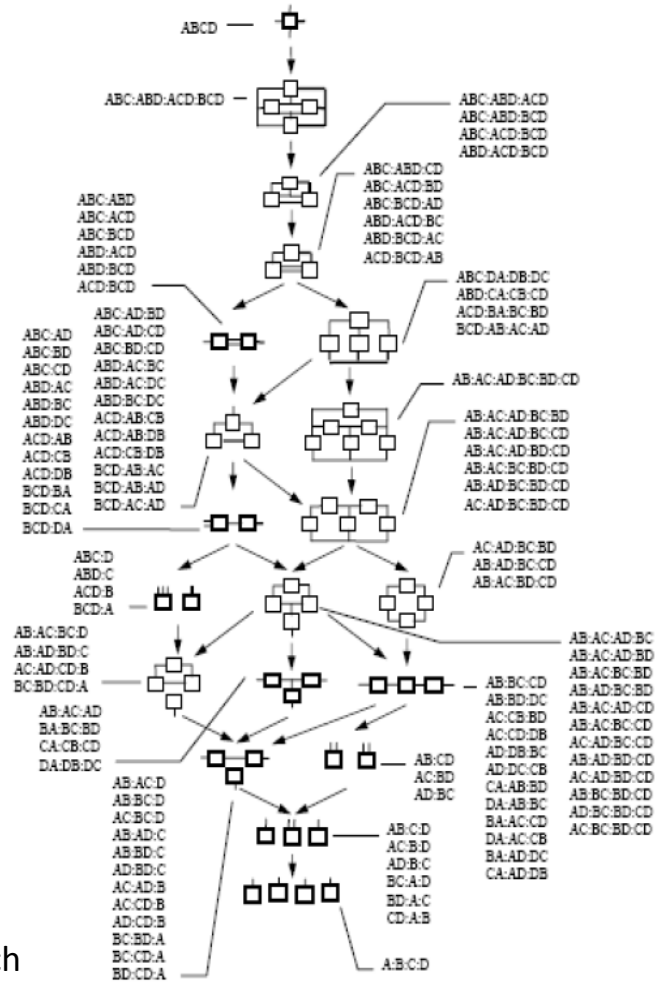
# RA & BN background continued

3 variable RA neutral system



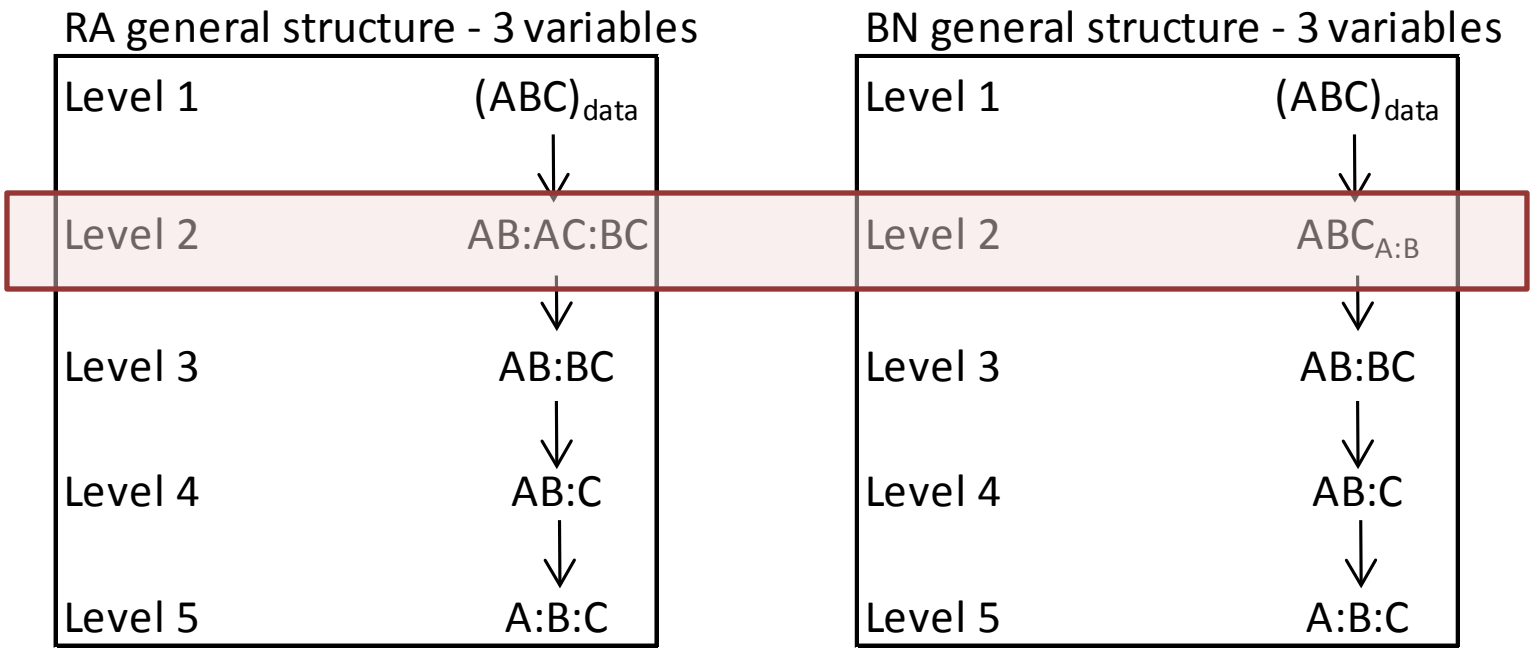

| RA Neutral System | # of specific structures |
|---|---|
| 3 Variable | 9 |
| 4 Variable | 114 |
| 6 Variable (not pictured) | 7,785,062 |

4 variable RA neutral system




*RA uses a beam search to explore the lattice, BN a greedy search

# RA & BN background continued




- The graphic above shows the lattice of structures for a 3 variable neutral RA system with loops compared to that of a 3 variable BN system, where A & B are IVs and C is a DV
- Lattices are the same except at Level 2, where
  - RA structure contains all permutations of dyadic relations for all 3 variables, whereas the BN structure preserves the triadic ABC relation & imposes independence between the two IVs

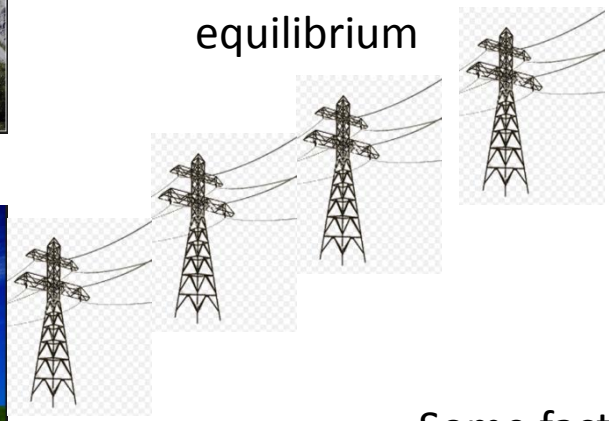
# RA and BN applied to the Power grid

- Electricity supply and demand is balanced instantaneously and continuously
- Many hundreds/thousands of variables potentially impact the supply/demand balance of electricity
- RA and BN were fed 31 potential explanatory variables to explore the lattice of possible structures to find the best model to predict supply/demand imbalance
- Compared results of the best RA and BN models to each other, also to linear regression

# Basic model structure

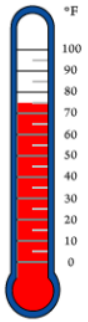
Electricity Supply

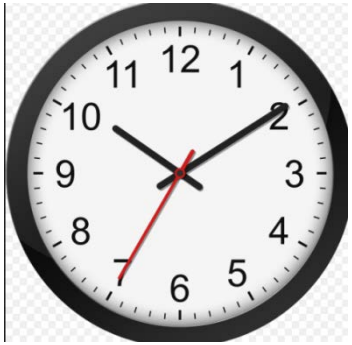
equilibrium



Electricity Demand (Load)

Some factors impacting supply/demand



# Preliminary modeling results - variables

- The table to the right shows a sample of explanatory/ independent variables & the response/dependent variable
- 31 Independent variables used to predict the response variable (DV) 'Balancing Reserves'
- Preliminary dataset randomly separated into training and test, approximately 500k samples in each
- Continuous independent variables were binned to 6 and 3 bins
- Response variable binned to 6 bins
- Variables included system signals such as expected demand and supply and environmental data such as temperature and wind speed.

Table: Sample of variables and summary statistics

| All Variables | | | | | |
|---|---|---|---|---|---|
| IV/DV | Variable Name | Variable Short Name | Cardinality | %ΔH(DV) | %Correct Training |
| Independent Variables | Wind Error | We6 | 6 | 14.7% | 37% |
| | Wind Error | We3 | 3 | 11.1% | 33% |
| | Wind Generation | Wg6 | 6 | 3.8% | 25% |
| | Wind Generation | Wg3 | 3 | 3.3% | 25% |
| | Wind Forecast | Wf6 | 6 | 3.0% | 24% |
| | Wind Forecast | Wf3 | 3 | 2.7% | 24% |
| | Load Error | Br6 | 6 | 0.8% | 20% |
| | Load Error | Tg6 | 6 | 0.6% | 20% |
| | Total Generation | Tg3 | 3 | 0.6% | 20% |
| | Total Generation | Mxt6 | 6 | 0.4% | 20% |
| | Max Temp. | Lfp6 | 6 | 0.4% | 19% |
| | Load Forecast | T6 | 6 | 0.3% | 19% |
| | Hydro Generation | H6 | 6 | 0.3% | 19% |
| | Thermal Generation | T3 | 3 | 0.3% | 19% |
| | Wind State | Ws | 3 | 0.2% | 19% |
| | Min Temp. | Mnt6 | 6 | 0.2% | 18% |
| | Load Forecast | Lfp3 | 3 | 0.2% | 19% |
| | Max Temp. | Mxt3 | 3 | 0.2% | 19% |
| | Load | L6 | 6 | 0.2% | 19% |
| | Hydro Generation | H3 | 3 | 0.2% | 18% |
| | Load Forecast | Lfn6 | 6 | 0.1% | 19% |
| | Load | L3 | 3 | 0.1% | 18% |
| | Min Temp. | Mnt3 | 3 | 0.1% | 18% |
| | Storm | S | 2 | 0.1% | 18% |
| | Inc Max | I6 | 6 | 0.1% | 18% |
| | Load Forecast | Lfn3 | 3 | 0.1% | 18% |
| | HLH/LLH | Hlh | 2 | 0.1% | 18% |
| | Dec Max | D6 | 6 | 0.1% | 18% |
| | Dec Max | D3 | 3 | 0.1% | 18% |
| | Inc Max | I3 | 3 | 0.0% | 18% |
| DV | Balancing Reserves | Br6 | 6 | 0.0% | 18% |

# Preliminary Results - RA

- Best RA model selected using incremental significance test
- Best model includes 7 of 31 possible explanatory variables
- Produces a %Correct of 42% overall, with highest percent correct in tails of distribution (State 1 and State 6)

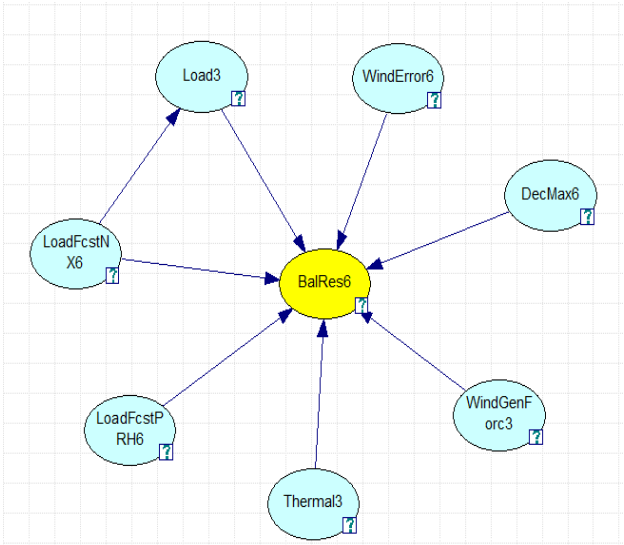
| Best RA Model no loops<br>IV: Hlh S Wf3 H3 We6 Mxt3 Le6 Br6 | | | |
|---|---|---|---|
| | %C | # C | # |
| **Total** | **42%** | **88,802** | **209,905** |
| State 1 | 72% | 23,206 | 32,289 |
| State 2 | 27% | 9,028 | 33,198 |
| State 3 | 26% | 9,248 | 35,591 |
| State 4 | 29% | 10,625 | 36,389 |
| State 5 | 41% | 15,112 | 37,002 |
| State 6 | 61% | 21,583 | 35,436 |

*Note 1: For each of the tables with % C results (that is %Correct) , States 1 - 6 represent low to increasingly higher amounts of the DV, operating reserves. This is important because state 1 represents the tail of the distribution of operating reserves on the low end and state 6 represents the tail of the distribution of operating reserves on the high end. These two states are of highest importance, that is, prediction accuracy in the tails is much more important than at the mean. Also, for clarity, #C in the table represents the number of correct predictions and the "#" column represents the total number predicted, therefore #C/# = %C.*

# Preliminary Results - BN

- Best BN model includes 7 of 31 possible explanatory variables
- Produces a %Correct of 45% overall, with highest percent correct in tails of distribution (State 1 and State 6)
- Uses more degrees freedom than RA, 7 variables have higher cardinality than the 7 variables in best RA model

| Greedy Thick Thinning All IVs | | | |
|---|---|---|---|
| | %C | # C | # |
| **Total** | **45%** | **95,290** | **209,905** |
| State 1 | 75% | 24,134 | 32,289 |
| State 2 | 33% | 10,985 | 33,198 |
| State 3 | 30% | 10,648 | 35,591 |
| State 4 | 29% | 10,621 | 36,389 |
| State 5 | 42% | 15,468 | 37,002 |
| State 6 | 66% | 23,434 | 35,436 |




*Note 1: For each of the tables with % C results (that is %Correct) , States 1 - 6 represent low to increasingly higher amounts of the DV, operating reserves. This is important because state 1 represents the tail of the distribution of operating reserves on the low end and state 6 represents the tail of the distribution of operating reserves on the high end. These two states are of highest importance, that is, prediction accuracy in the tails is much more important than at the mean. Also, for clarity, #C in the table represents the number of correct predictions and the "#" column represents the total number predicted, therefore #C/# = %C.*

# Preliminary Results - Linear Regression

- A linear regression model was developed using the same independent variables found in the best RA model
- Preliminary results indicate that regression underperforms both RA and BN

Results of linear regression model

| Linear Regression with RA IVs | | | |
|---|---|---|---|
| | %C | # C | # |
| **Total** | **37%** | **76,654** | **209,905** |
| State 1 | 53% | 17,084 | 32,289 |
| State 2 | 28% | 9,341 | 33,198 |
| State 3 | 33% | 11,869 | 35,591 |
| State 4 | 43% | 15,558 | 36,389 |
| State 5 | 25% | 9,213 | 37,002 |
| State 6 | 38% | 13,589 | 35,436 |

# Summary

Table: Preliminary Modeling Results

| | RA | BN | Regression |
|---|---|---|---|
| Total % Correct | **42%** | **45%** | **37%** |
| State 1 | 72% | 75% | 53% |
| State 2 | 27% | 33% | 28% |
| State 3 | 26% | 30% | 33% |
| State 4 | 29% | 29% | 43% |
| State 5 | 41% | 42% | 25% |
| State 6 | 61% | 66% | 38% |

- Models of the electric grid were developed to compare RA, BN and linear regression
- Preliminary results show RA and BN models predict
  - (a) better on average
  - (b) much better in the distribution tails (States 1 & 6)
- Future extensions of this research include :
  - Building empirical models and testing results against the exploratory models shown here
  - Increasing the number of possible explanatory variables to include time series variables
  - Develop time series regression model as well as compare to other machine learning methods such as support vector machines and artificial neural networks