

# The Bayesys data and Bayesian Network repository

Anthony C. Constantinou<sup>1,2</sup>, Yang Liu<sup>1</sup>, Kiattikun Chobtham<sup>1</sup>, Zhigao Guo<sup>1</sup>, and Neville K. Kitson<sup>1</sup>

Version 1.02<sup>1</sup>

- i. [Bayesian Artificial Intelligence](#) research lab, Risk and Information Management (RIM) research group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK, E1 4FZ.  
E-mail: [a.constantinou@qmul.ac.uk](mailto:a.constantinou@qmul.ac.uk)
- ii. The Alan Turing Institute, UK, British Library, London, UK, NW1 2DB.



---

<sup>1</sup> **Citation:** Constantinou, A. C., Liu, Y., Chobtham, K., Guo, Z., and Kitson, N. K. (2020). The Bayesys data and Bayesian network repository. *Queen Mary University of London*, London, UK. [Online]. Available: <http://bayesian-ai.eecs.qmul.ac.uk/bayesys/> or <http://www.bayesys.com>

## Table of Contents

Copyright notice.....	3
Acknowledgements .....	4
Introduction.....	5
1. Case study: ASIA .....	6
2. Case study: ALARM.....	8
3. Case study: PATHFINDER .....	11
4. Case study: SPORTS .....	13
5. Case study: PROPERTY .....	15
6. Case study: FORMED .....	18
References.....	22

## Copyright notice



Copyright © Bayesys.com. This report on the data repository of Bayesys is distributed under the terms of a CC BY-SA license: [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

THE BAYESYS DATA AND BAYESIAN NETWORK REPOSITORY IS DISTRIBUTED AND LICENSED FREE OF CHARGE IN THE HOPE IT WILL BE USEFUL. BECAUSE OF THIS, THERE IS NO WARRANTY OF ANY KIND FOR THE ACCURACY OR USEFULNESS OF THIS INFORMATION EXCEPT AS REQUIRED BY APPLICABLE LAW OR EXPRESSLY AGREED IN WRITING.

## Acknowledgements

1. This work is supported by the ERSRC Fellowship project EP/S001646/1 “*Bayesian Artificial Intelligence for Decision Making under Uncertainty*” [1], by the project partner Agena Ltd, and by The Alan Turing Institute.

## Introduction

This document provides access to discrete data, graphs, and discrete Bayesian Networks (BNs) associated with various case studies intended for structure learning experiments with noisy data, and closely follows the work on the Bayesys system [2]. The material presented in this report is largely based on the work in [3]. Each of the sections that follow correspond to a particular case study.

# 1. Case study: ASIA

A small traditional toy network for diagnosing patients at a clinic [4]:

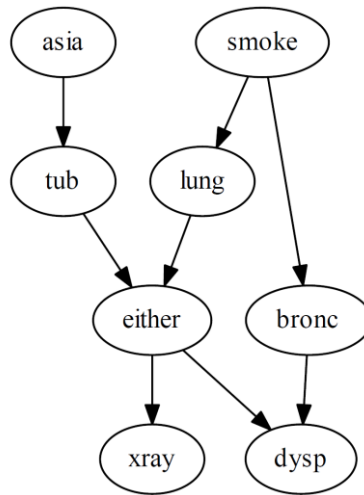
Number of variables:	8
Number of edges:	8
Number of free parameters:	18
Maximum in-degree:	2

The data available from this case study comes from the experiments in [3]. The following data are available:

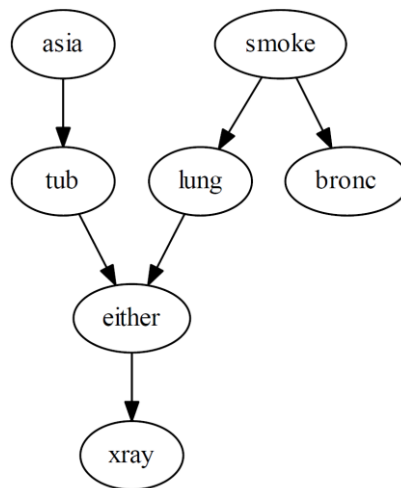
- The true graph shown in Fig 1.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 1.1; i.e., the “*DAG*” in column *True graph* for experiments 1, 2, 3, 4, 5, and 10.
- A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 1.1; i.e., the “*Link*” in column *Dataset*, associated with experiment 1.
- Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 1.1; i.e., the “*Link*” in column *Dataset*, for experiments 2 to 16.
- Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is a Maximal Ancestral Graph (MAG), rather than a Directed Acyclic Graph (DAG). For the ASIA case study, only one MAG is used, shown in Fig 1.2. You will find the link to download the true MAG in Table 1.1; i.e., “*MAG-10*” in column *True graph* for experiments 9, 12, 14, and 16.

**Table 1.1.** ASIA datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
6	S5	-	-
7	S10	-	-
8	L5	-	-
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 1.2)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 1.1)
11	cMS	-	-
12	cML	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 1.2)
13	cIS	-	-
14	cIL	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 1.2)
15	cSL	-	-
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 1.2)



**Fig 1.1.** The true graph of ASIA. Total variables: 8. Total edges: 8.



**Fig 1.2.** The true MAG-10 of ASIA. Total variables: 7. Total edges: 6. Latent variables: dysp.

## 2. Case study: ALARM

A medium traditional network based on an alarm message system for patient monitoring [5]:

Number of variables:	37
Number of edges:	46
Number of free parameters:	509
Maximum in-degree:	4

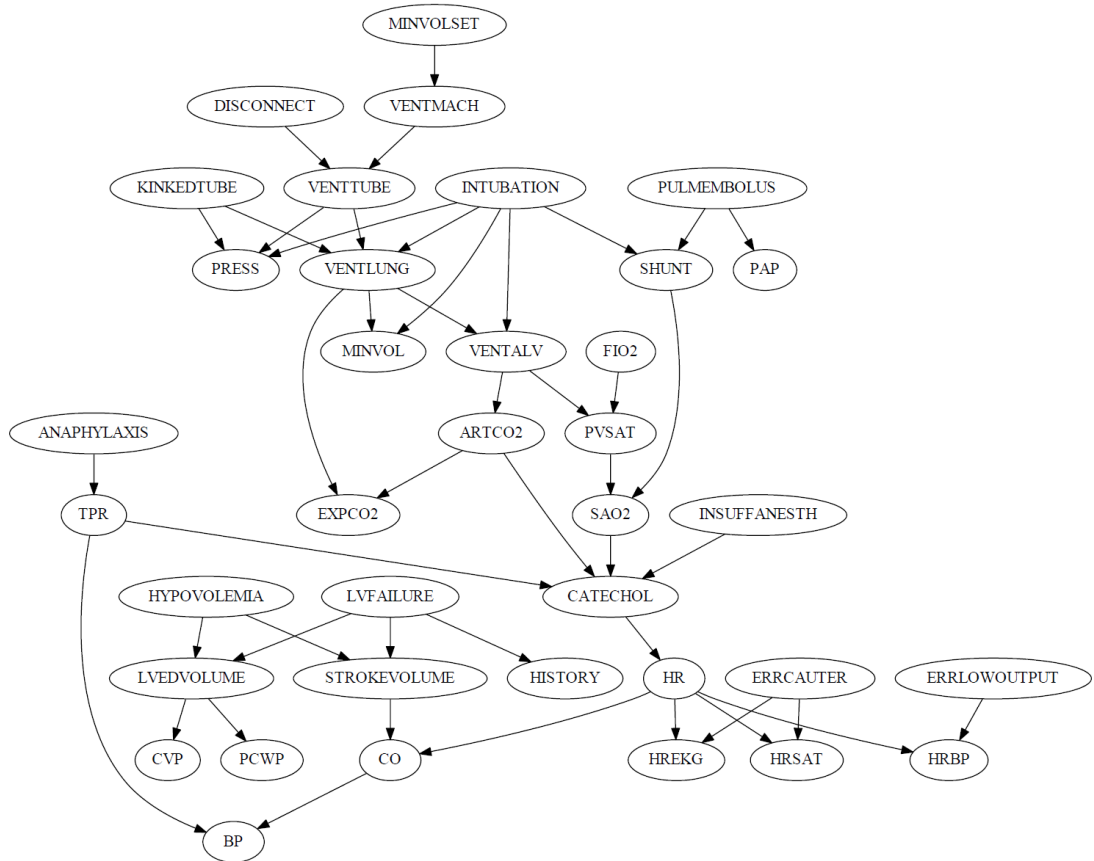
The data available from this case study comes from the experiments in [3]. The following data are available:

- The true graph shown in Fig 2.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 2.1; i.e., the “DAG” in column *True graph* for experiments 1, 2, 3, 4, 5, 6, 7, 10, 11, and 13.
- A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 2.1; i.e., the “Link” in column *Dataset*, associated with experiment 1.
- Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 2.1; i.e., the “Link” in column *Dataset*, for experiments 2 to 16.
- Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is MAG, rather than a DAG. For the ALARM case study, two MAGs are used, shown in Figs 2.2 and 2.3. You will find the links to download the true MAGs in Table 2.1; i.e., “MAG-5” and “MAG-10” in column *True graph* for experiments 8, 9, 12, 14, 15, and 16.

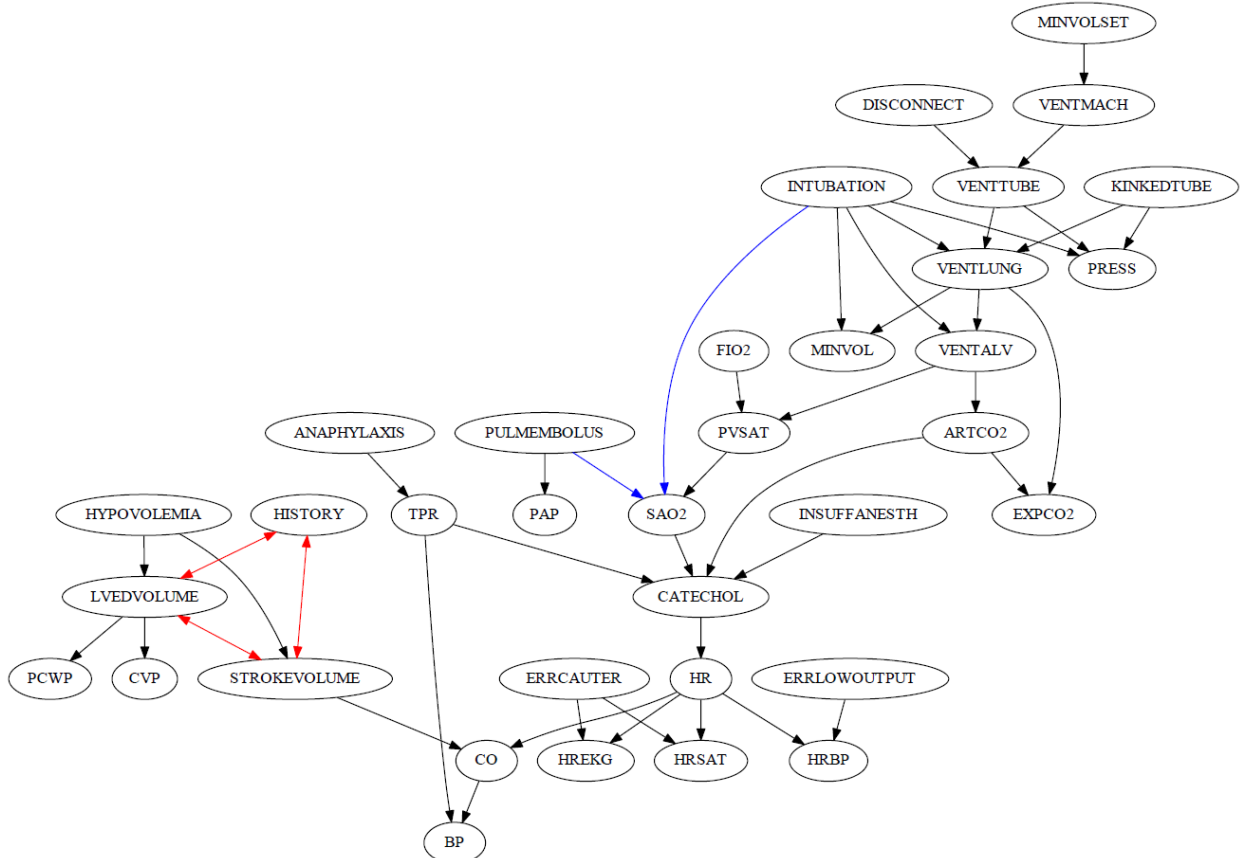
**Table 2.1.** ALARM datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
6	S5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
7	S10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
8	L5	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 2.2)
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 2.3)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
11	cMS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
12	cML	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 2.2)
13	cIS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 2.1)
14	cIL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 2.2)
15	cSL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 2.2)
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 2.2)

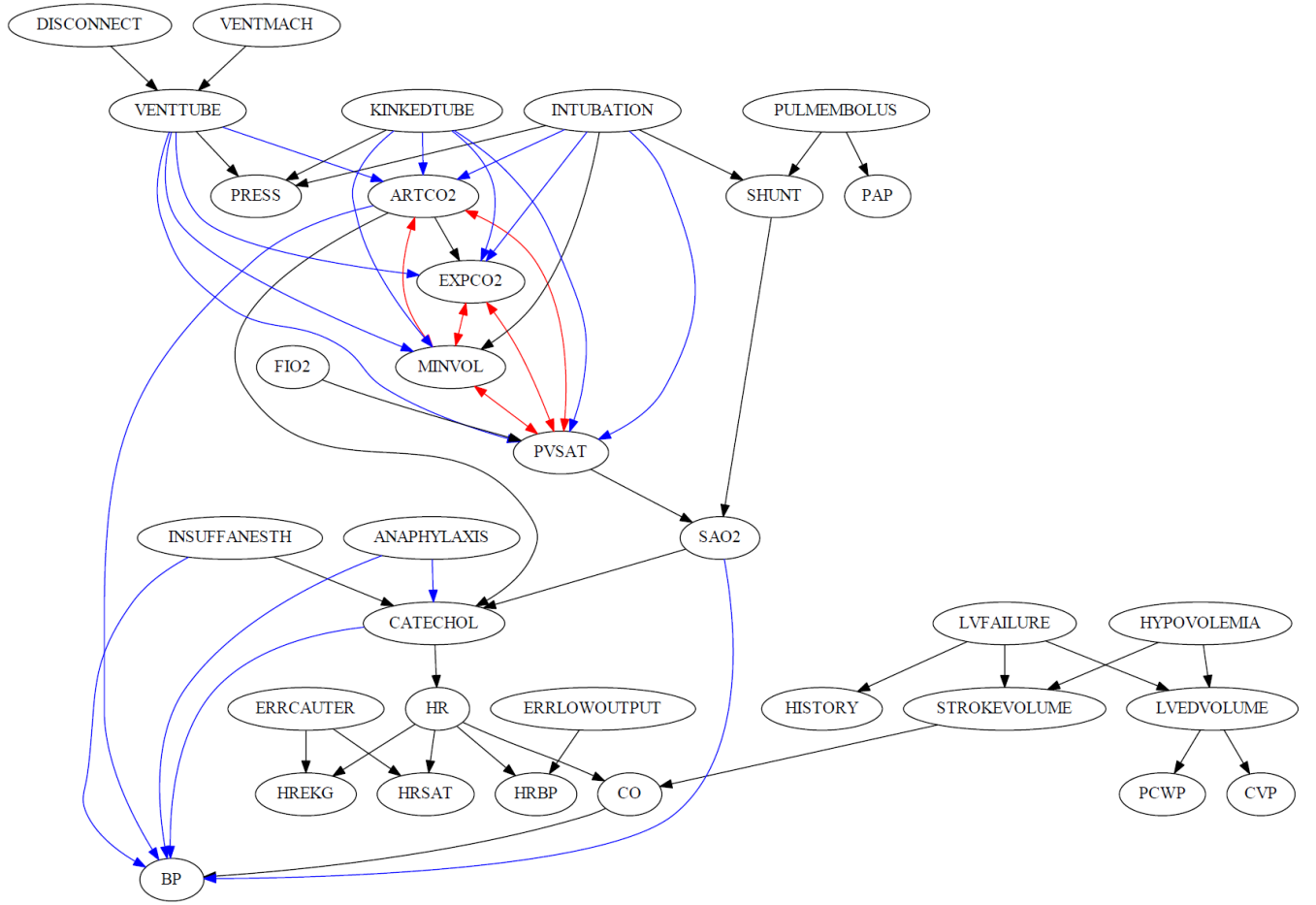




**Fig 2.1.** The true graph of ALARM. Total variables: 37. Total edges: 46.



**Fig 2.2.** The true MAG-5 of ALARM. Total variables: 35. Total edges: 46. Latent variables: LVFAILURE, SHUNT. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG.



**Fig 2.3.** The true MAG-10 of ALARM. Total variables: 33. Total edges: 55. Latent variables: VENTLUNG, MINVOLSET, VENTALV, TPR. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG.

### 3. Case study: PATHFINDER

A very large traditional network, though less popular than the previous two, that was designed to assist surgical pathologists with the diagnosis of lymph-node diseases [6]:

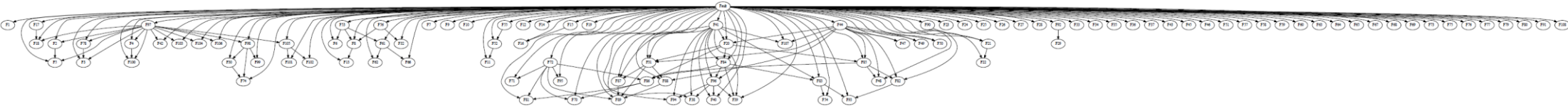
Number of variables:	109
Number of edges:	195
Number of free parameters:	71890
Maximum in-degree:	5

The data available from this case study comes from the experiments in [3]. The following data are available:

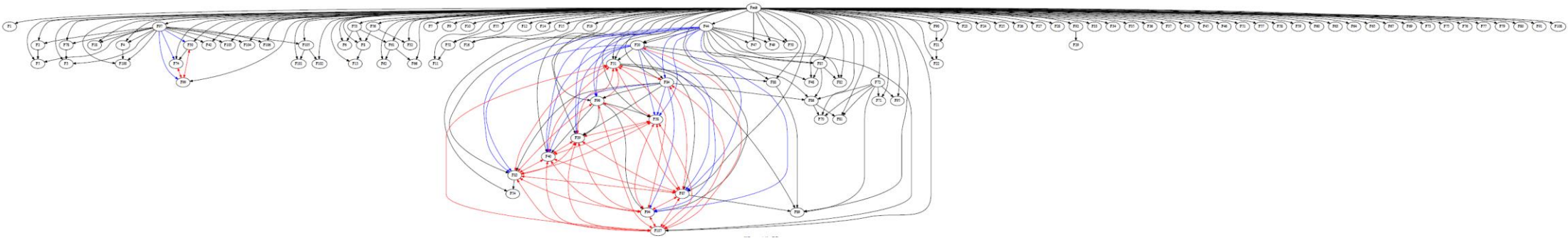
- The true graph shown in Fig 3.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 3.1; i.e., the “*DAG*” in column *True graph* for experiments 1, 2, 3, 4, 5, 6, 7, 10, 11, and 13.
- A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 3.1; i.e., the “*Link*” in column *Dataset*, associated with experiment 1.
- Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 3.1; i.e., the “*Link*” in column *Dataset*, for experiments 2 to 16.
- Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is MAG, rather than a DAG. For the PATHFINDER case study, two MAGs are used, shown in Figs 3.2 and 3.3. You will find the links to download the true MAGs in Table 3.1; i.e., “*MAG-5*” and “*MAG-10*” in column *True graph* for experiments 8, 9, 12, 14, 15, and 16.

**Table 3.1.** PATHFINDER datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

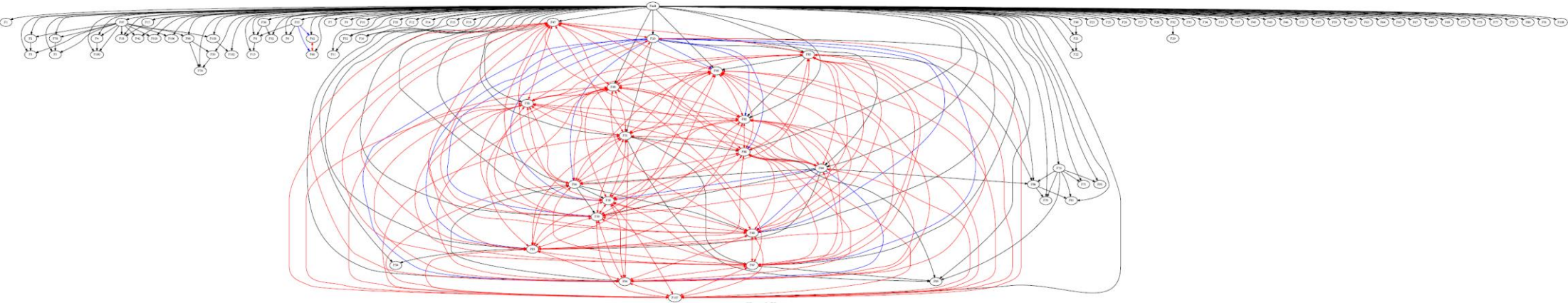
Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
6	S5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
7	S10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
8	L5	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 3.2)
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 3.3)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
11	cMS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
12	cML	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 3.2)
13	cIS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 3.1)
14	cIL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 3.2)
15	cSL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 3.2)
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 3.2)



**Fig 3.1.** The true graph of PATHFINDER. Total variables: 109. Total edges: 195.



**Fig 3.2.** The true MAG-5 of PATHFINDER. Total variables: 104. Total edges: 230. Latent variables: F17, F41, F98, F68, F93. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG.



**Fig 3.3.** The true MAG-10 of PATHFINDER. Total variables: 99. Total edges: 297. Latent variables: F41, F44, F24, F36, F58, F61, F76, F93, F99, F101, F104. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG.

## 4. Case study: SPORTS

A small BN that combines football team ratings with various team performance statistics to predict various match score outcomes [7]:

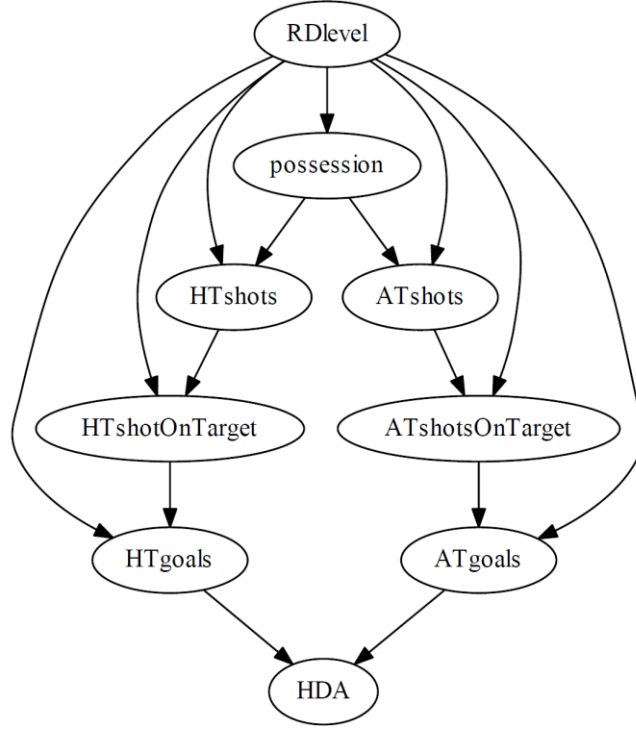
Number of variables:	9
Number of edges:	15
Number of free parameters:	1049
Maximum in-degree:	2

The data available from this case study comes from the experiments in [3]. The following data are available:

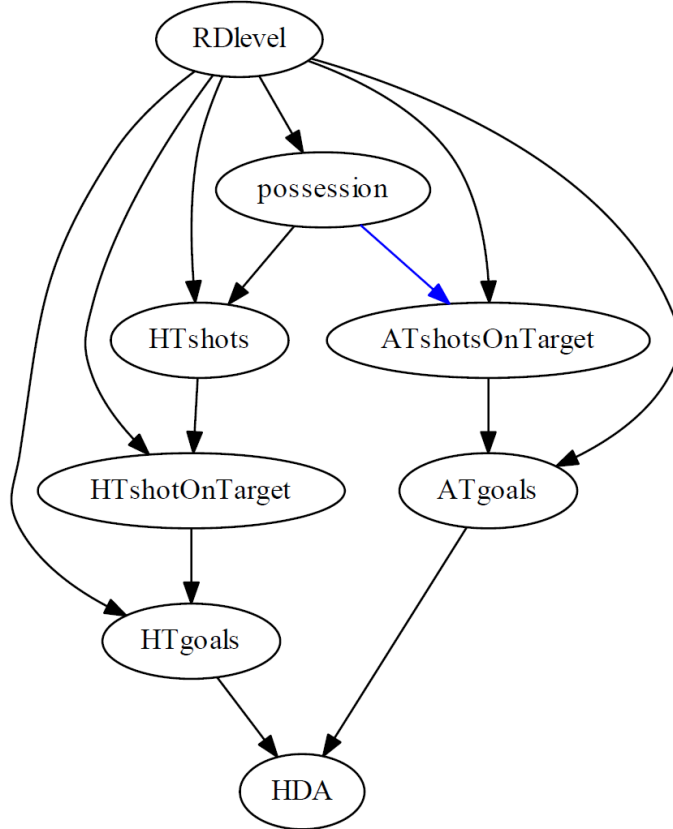
- The true graph shown in Fig 4.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 4.1; i.e., the “*DAG*” in column *True graph* for experiments 1, 2, 3, 4, 5, 7, 10, 11, and 13.
- A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 4.1; i.e., the “*Link*” in column *Dataset*, associated with experiment 1.
- Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 4.1; i.e., the “*Link*” in column *Dataset*, for experiments 2 to 16.
- Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is MAG, rather than a DAG. For the SPORTS case study, one MAGs is used, shown in Fig 2.2. You will find the link to download the true MAG in Table 4.1; i.e., “*MAG-10*” in column *True graph* for experiments 9, 12, 14, 15, and 16.

**Table 4.1.** SPORTS datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
6	S5	-	-
7	S10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
8	L5	-	-
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 4.2)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
11	cMS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
12	cML	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 4.2)
13	cIS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 4.1)
14	cIL	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 4.2)
15	cSL	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 4.2)
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 4.2)



**Fig 4.1.** The true graph of SPORTS. Total variables: 9. Total edges: 15.



**Fig 4.2.** The true MAG-10 of SPORTS. Total variables: 8. Total edges: 13. Latent variables: ATshots. Blue edges represent arcs in MAG that are not present in the ground truth DAG.



## 5. Case study: PROPERTY

A medium BN that assesses investment decisions in the UK property market [8]:

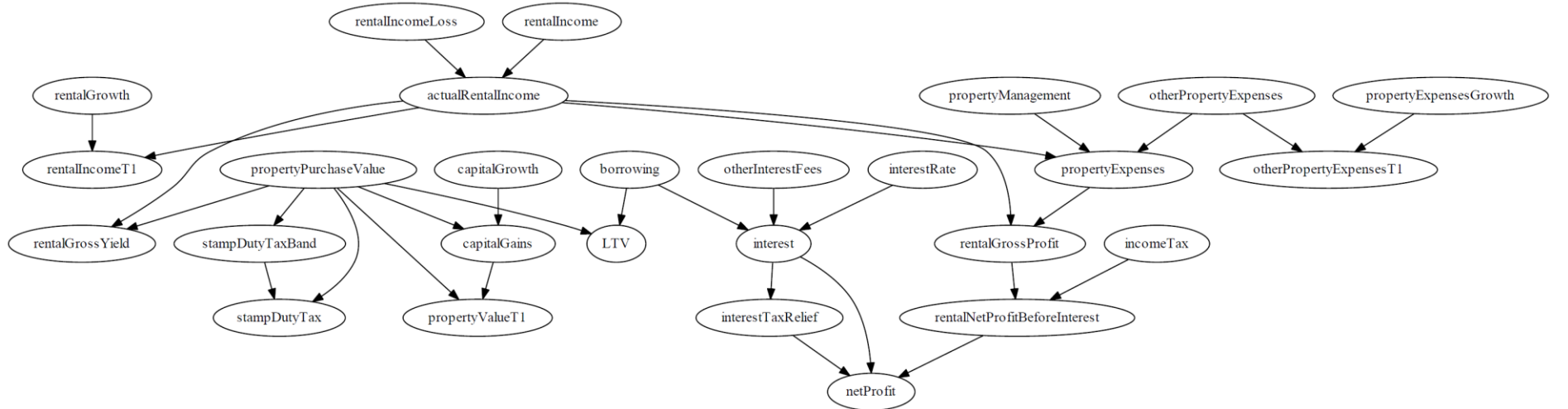
Number of variables:	27
Number of edges:	31
Number of free parameters:	3056
Maximum in-degree:	3

The data available from this case study comes from the experiments in [3]. The following data are available:

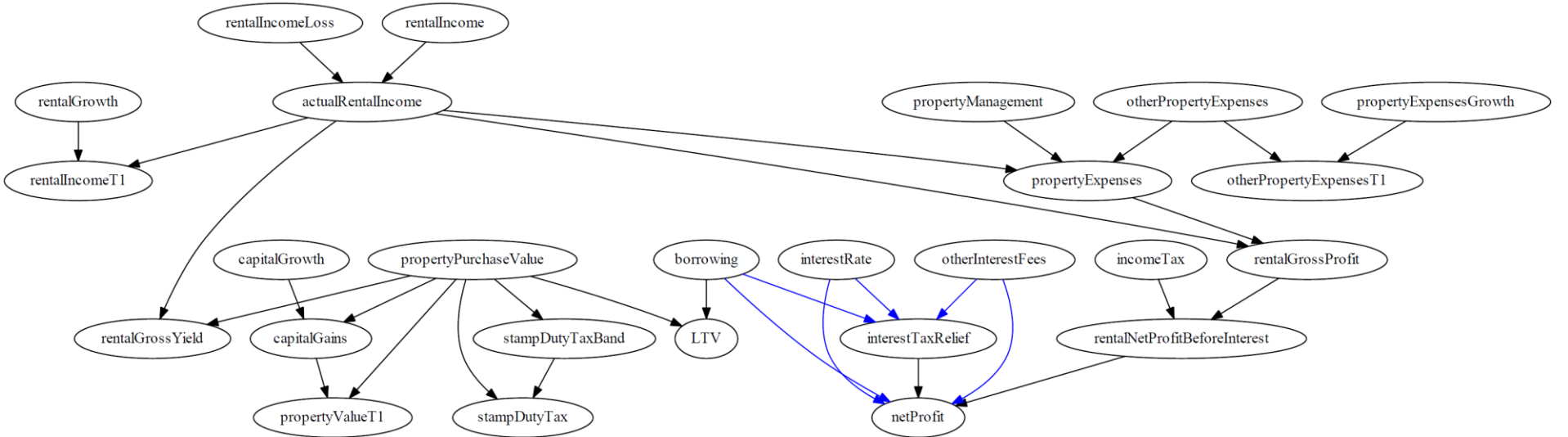
- e) The true graph shown in Fig 5.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 5.1; i.e., the “*DAG*” in column *True graph* for experiments 1, 2, 3, 4, 5, 6, 7, 10, 11, and 13.
- f) A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 5.1; i.e., the “*Link*” in column *Dataset*, associated with experiment 1.
- g) Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 5.1; i.e., the “*Link*” in column *Dataset*, for experiments 2 to 16.
- h) Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is MAG, rather than a DAG. For the PROPERTY case study, two MAGs are used, shown in Figs 5.2 and 5.3. You will find the links to download the true MAGs in Table 5.1; i.e., “*MAG-5*” and “*MAG-10*” in column *True graph* for experiments 8, 9, 12, 14, 15, and 16.

**Table 5.1.** PROPERTY datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
6	S5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
7	S10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
8	L5	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 5.2)
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 5.3)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
11	cMS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
12	cML	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 5.2)
13	cIS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 5.1)
14	cIL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 5.2)
15	cSL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 5.2)
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 5.2)

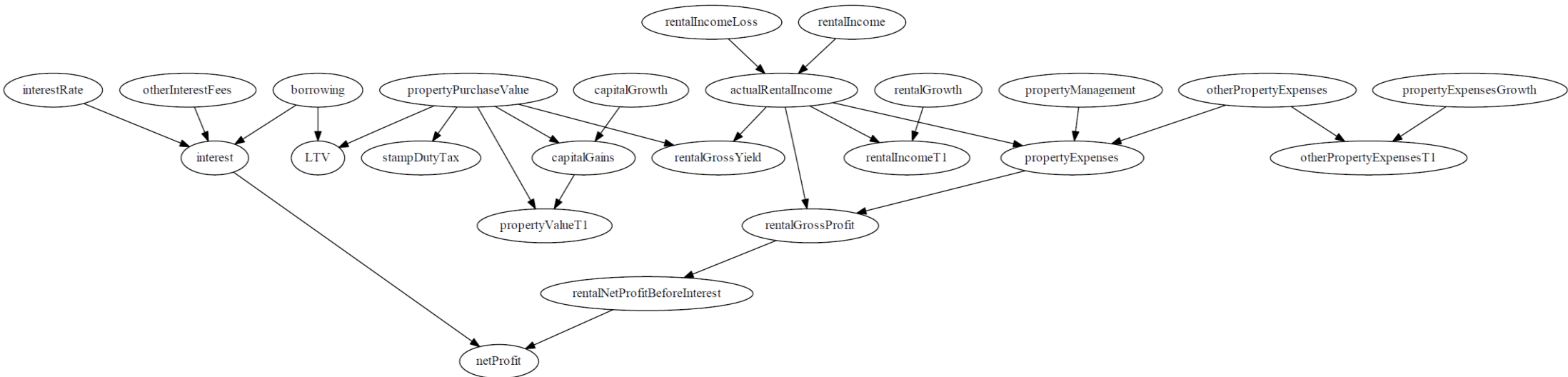


**Fig 5.1.** The true graph of PROPERTY. Total variables: 27. Total edges: 31.



**Fig 5.2.** The true MAG-5 of PROPERTY. Total variables: 26. Total edges: 32. Latent variables: Interest. Blue edges represent arcs in MAG that are not present in the ground truth DAG.





**Fig 5.3.** The true MAG-10 of PROPERTY. Total variables: 24. Total edges: 26. Latent variables: stampDutyTaxBand, incomeTax, interestTaxRelief.

## 6. Case study: FORMED

A large BN that captures the risk of violent reoffending of mentally ill prisoners, along with multiple interventions for managing this risk [9]:

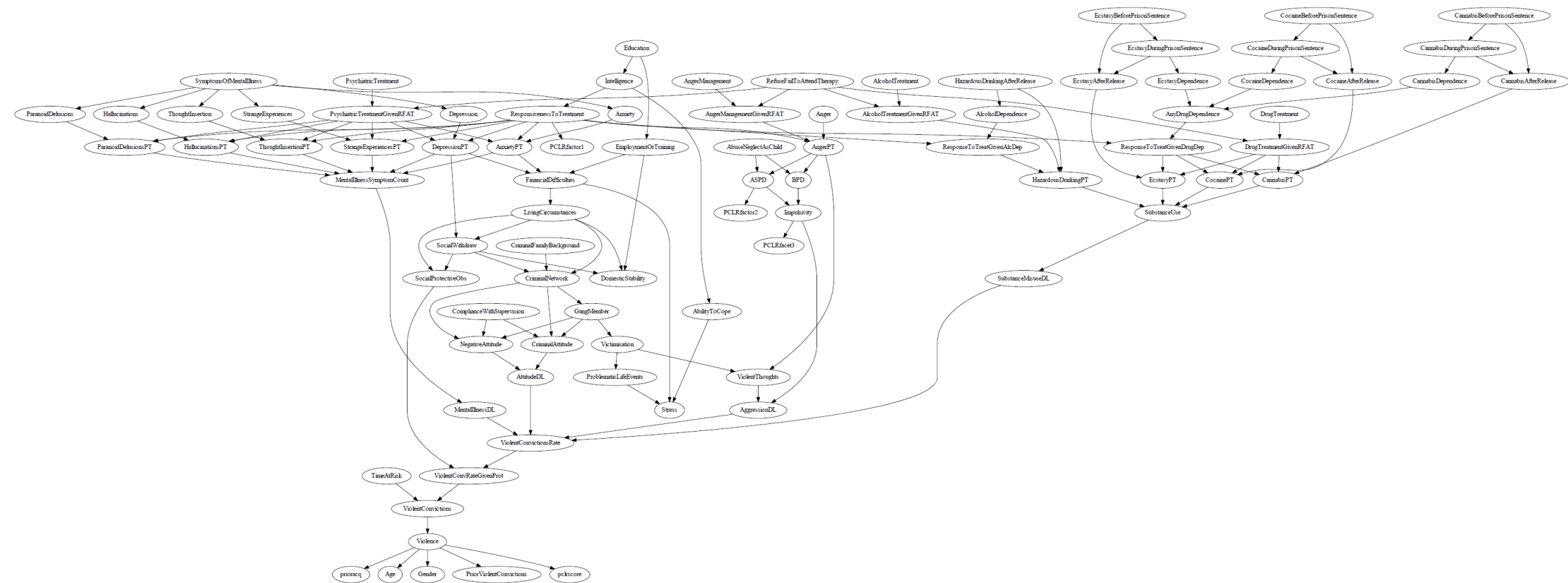
Number of variables:	88
Number of edges:	138
Number of free parameters:	912
Maximum in-degree:	6

The data available from this case study comes from the experiments in [3]. The following data are available:

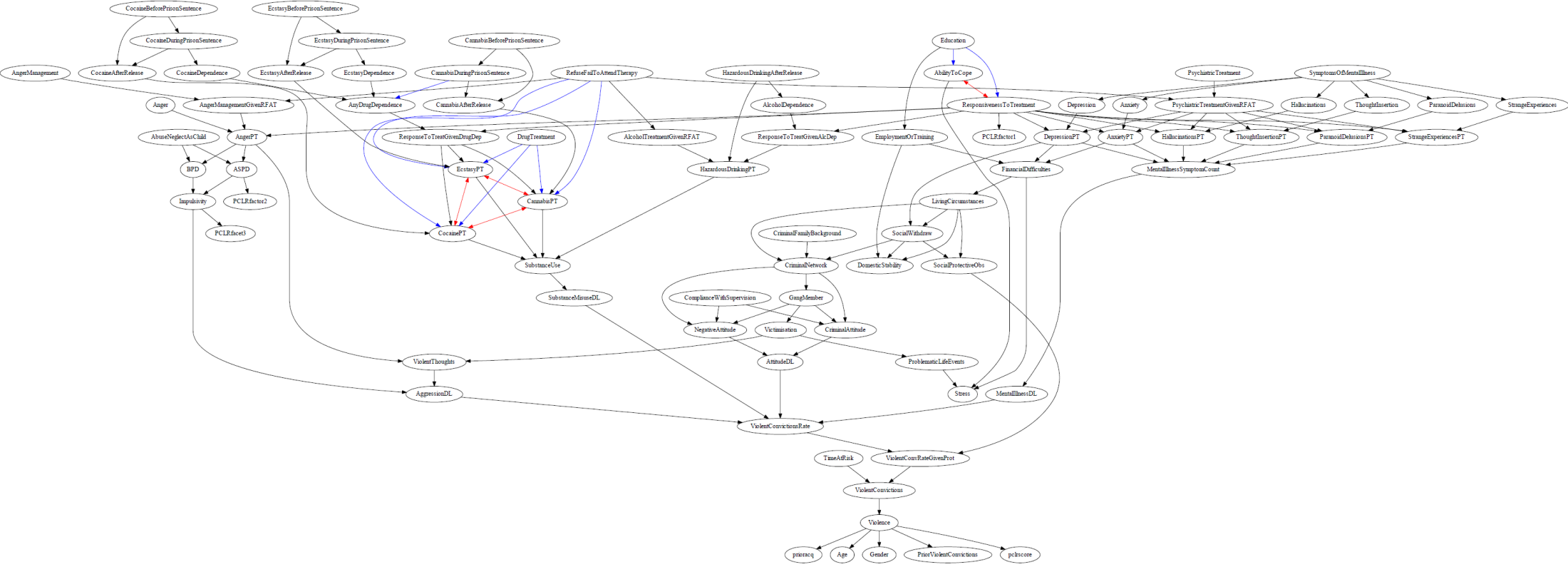
- The true graph shown in Fig 6.1 which can be downloaded as a CSV file. You will find the link to download this file in Table 6.1; i.e., the “*DAG*” in column *True graph* for experiments 1, 2, 3, 4, 5, 6, 7, 10, 11, and 13.
- A synthetic dataset that has been generated from the true graph. The dataset is available to download in five different sample sizes: 0.1k, 1k, 10k, 100k, and 1000k samples. You will find the link to download this file in Table 6.1; i.e., the “*Link*” in column *Dataset*, associated with experiment 1.
- Another 15 synthetic datasets that have been generated from the true graph, but with different types of noise incorporated into the data. The types of noise are covered in [3]. You will find the links to download these files in Table 6.1; i.e., the “*Link*” in column *Dataset*, for experiments 2 to 16.
- Some of the types of noise discussed in (c) above involve latent variables. Those cases assume that the true graph is MAG, rather than a DAG. For the FORMED case study, two MAGs are used, shown in Figs 6.2 and 6.3. You will find the links to download the true MAGs in Table 6.1; i.e., “*MAG-5*” and “*MAG-10*” in column *True graph* for experiments 8, 9, 12, 14, 15, and 16.

**Table 6.1.** FORMED datasets and graphs where *N* indicates *No noise*, *M* indicates *Missing values*, *I* indicates *Incorrect values*, *S* indicates *Merged states*, *L* indicates *Latent variables*, *c* indicates a *Combo* case that incorporates multiple types of noise, and numerical values 5 and 10 indicate the rate of noise for each data point. Each experiment consists of five different sample sizes; from 0.1k to 1000k samples. For details refer to [3].

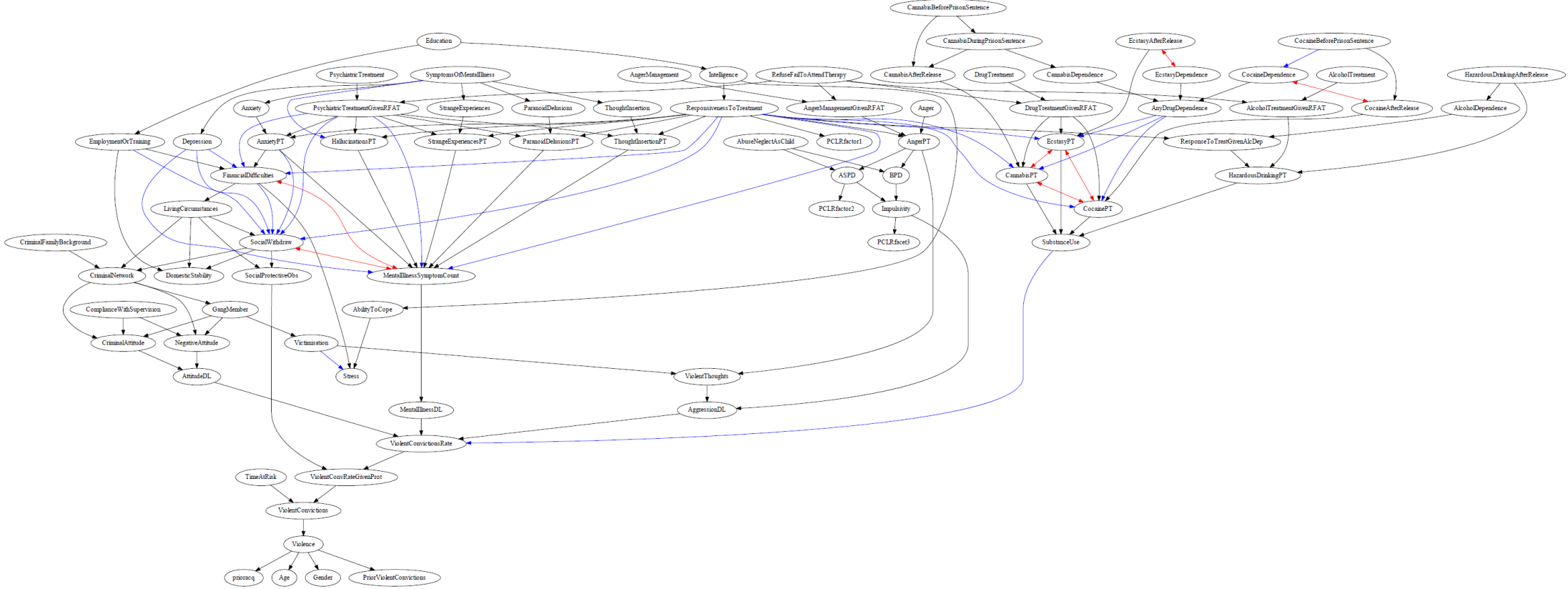
Experiment no.	Experiment	Dataset (.csv)	True graph (.csv)
1	N	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
2	M5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
3	M10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
4	I5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
5	I10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
6	S5	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
7	S10	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
8	L5	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 6.2)
9	L10	<a href="#">Link</a>	<a href="#">MAG-10</a> (Fig 6.3)
10	cMI	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
11	cMS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
12	cML	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 6.2)
13	cIS	<a href="#">Link</a>	<a href="#">DAG</a> (Fig 6.1)
14	cIL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 6.2)
15	cSL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 6.2)
16	cMISL	<a href="#">Link</a>	<a href="#">MAG-5</a> (Fig 6.2)



**Fig 6.1.** The true graph of FORMED. Total variables: 88. Total edges: 138.



**Fig 6.2.** The true MAG-5 of FORMED. Total variables: 84. Total edges: 140. Latent variables: Intelligence, CannabisDependence, AlcoholTreatment, DrugTreatmentGivenRFAT. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG.



**Fig 6.3.** The true MAG-10 of FORMED. Total variables: 79. Total edges: 142. Latent variables: ProblematicLifeEvents, pclr score, EcstasyBeforePrisonSentence, EcstasyDuringPrisonSentence, CocaineDuringPrisonSentence, SubstanceMisuseDL, ResponseToTreatGivenDrugDep, DepressionPT, Hallucinations. Blue and red edges represent arcs and bi-directed edges in MAG, respectively, that are not present in the ground truth DAG

## References

- [1] Constantinou, A. C. (2018). Bayesian Artificial Intelligence for Decision Making under Uncertainty. *Engineering and Physical Sciences Research Council (EPSRC)*, EP/S001646/1.
- [2] Constantinou, A. (2019). The Bayesys user manual. *Queen Mary University of London*, London, UK. [Online]. Available: <http://bayesian-ai.eecs.qmul.ac.uk/bayesys/> and <http://www.bayesys.com>
- [3] Constantinou, A. C., Liu, Y., Chobtham, K., Guo, Z., and Kitson, N. K. (2020). Large-scale empirical validation of Bayesian Network structure learning algorithms with noisy data. [arXiv:2005.09020](https://arxiv.org/abs/2005.09020) [cs.LG]
- [4] Lauritzen, S., and Spiegelhalter, D. (1988). Local Computation with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 50, Iss., pp. 157–224.
- [5] Beinlich, I. A., Suermondt, H. J., Chavez, R. M., and Cooper, G. F. (1989). The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks. In *Proceedings of the 2nd European Conference on Artificial Intelligence in Medicine*, pp. 247–256.
- [6] Heckerman, D., Horwitz, E., Nathwani, B. (1992). Towards Normative Expert Systems: Part I. The Pathfinder Project. *Methods of Information in Medicine*, Vol. 31, pp. 90–105.
- [7] Constantinou, A. C. (2019). Asian handicap football betting with rating-based hybrid Bayesian networks. [arXiv:2003.09384](https://arxiv.org/abs/2003.09384) [stat.AP]
- [8] Constantinou, A. C., and Fenton, N. (2017) The future of the London Buy-To-Let property market: Simulation with Temporal Bayesian Networks. *PLoS ONE*, Vol. 12, Iss. 6, e0179297.
- [9] Constantinou, A. C., Freestone, M., Marsh, W., Fenton, N., and Coid, J. (2015). Risk assessment and risk management of violent reoffending among prisoners. *Expert Systems with Applications*, Vol. 42, Iss. 21, pp. 7511–7529.