Intersection-Validation: A Method for Evaluating Structure Learning without Ground Truth

Supplement

Jussi Viinikka, Ralf Eggeling, Mikko Koivisto
Department of Computer Science, University of Helsinki, Finland
{jussivii,eggeling,mkhkoivi}@cs.helsinki.fi

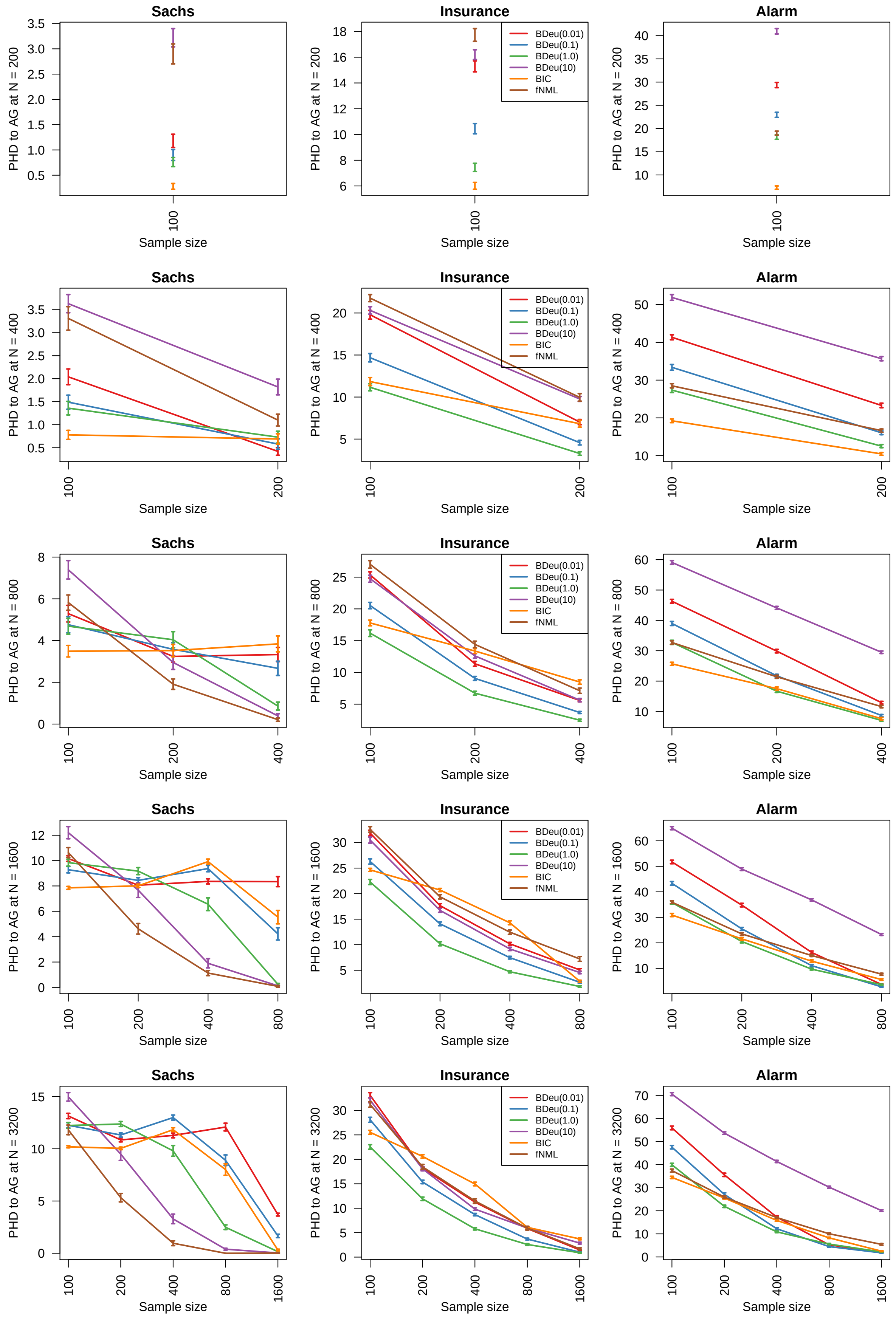
Contents

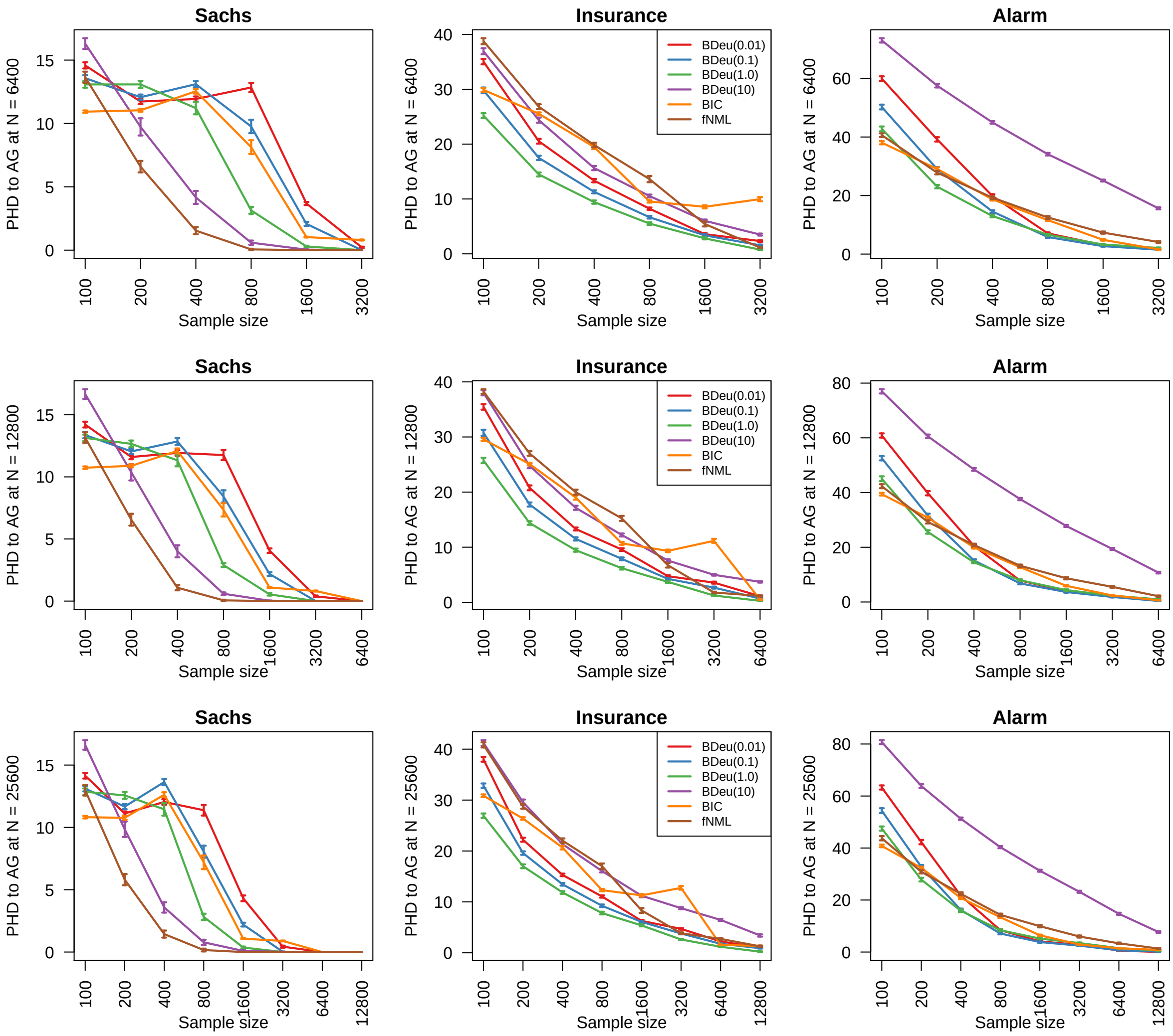
1 Partial Hamming distances **2**

2 Ground truth DAGs, CPDAG, and agreement graphs **4**
2.1 Alarm . . . 4
2.2 Insurance . . . 6
2.3 Sachs . . . 8

3 Performance of intersection-validation in pairwise comparison **10**

1 Partial Hamming distances

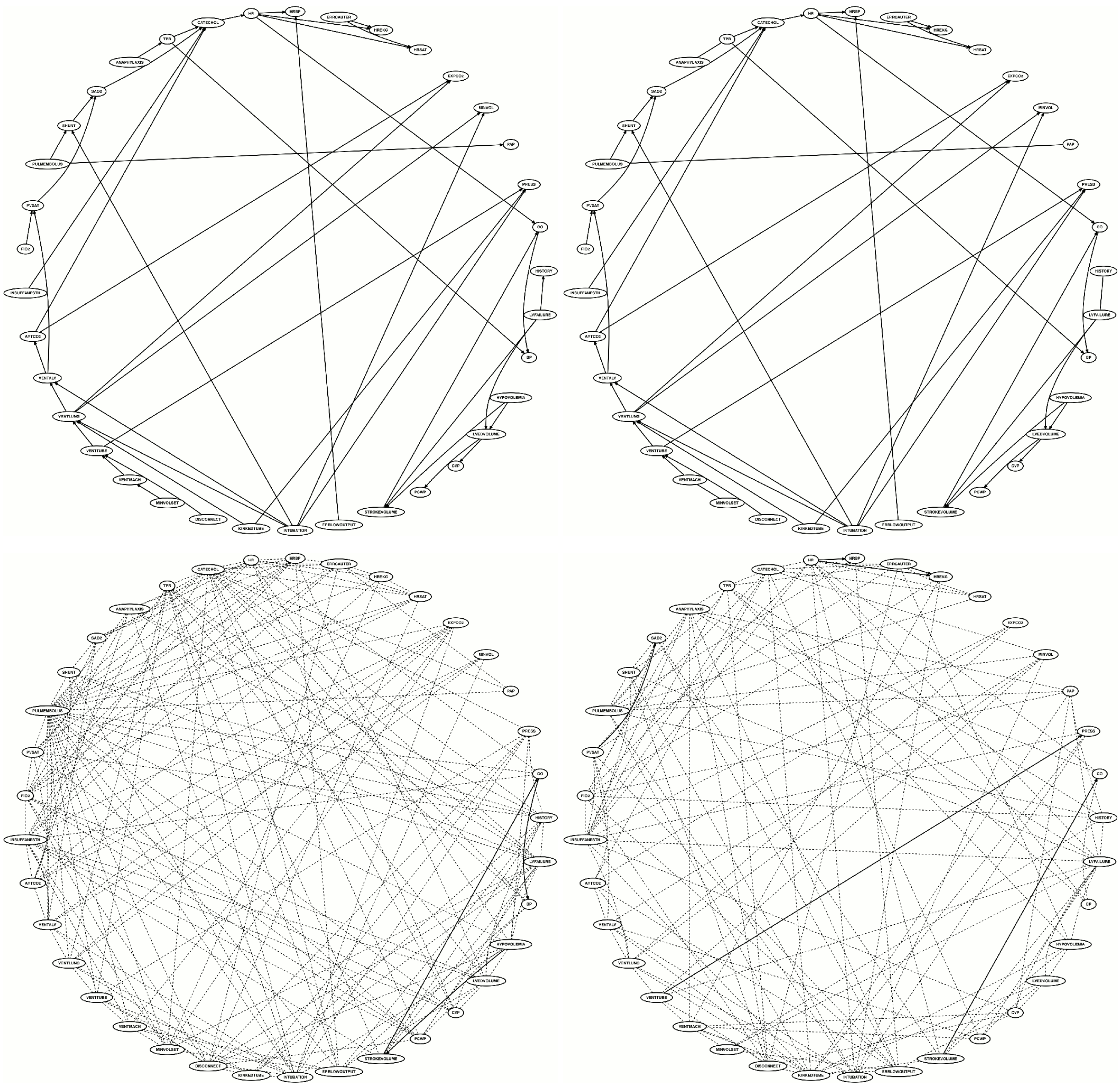


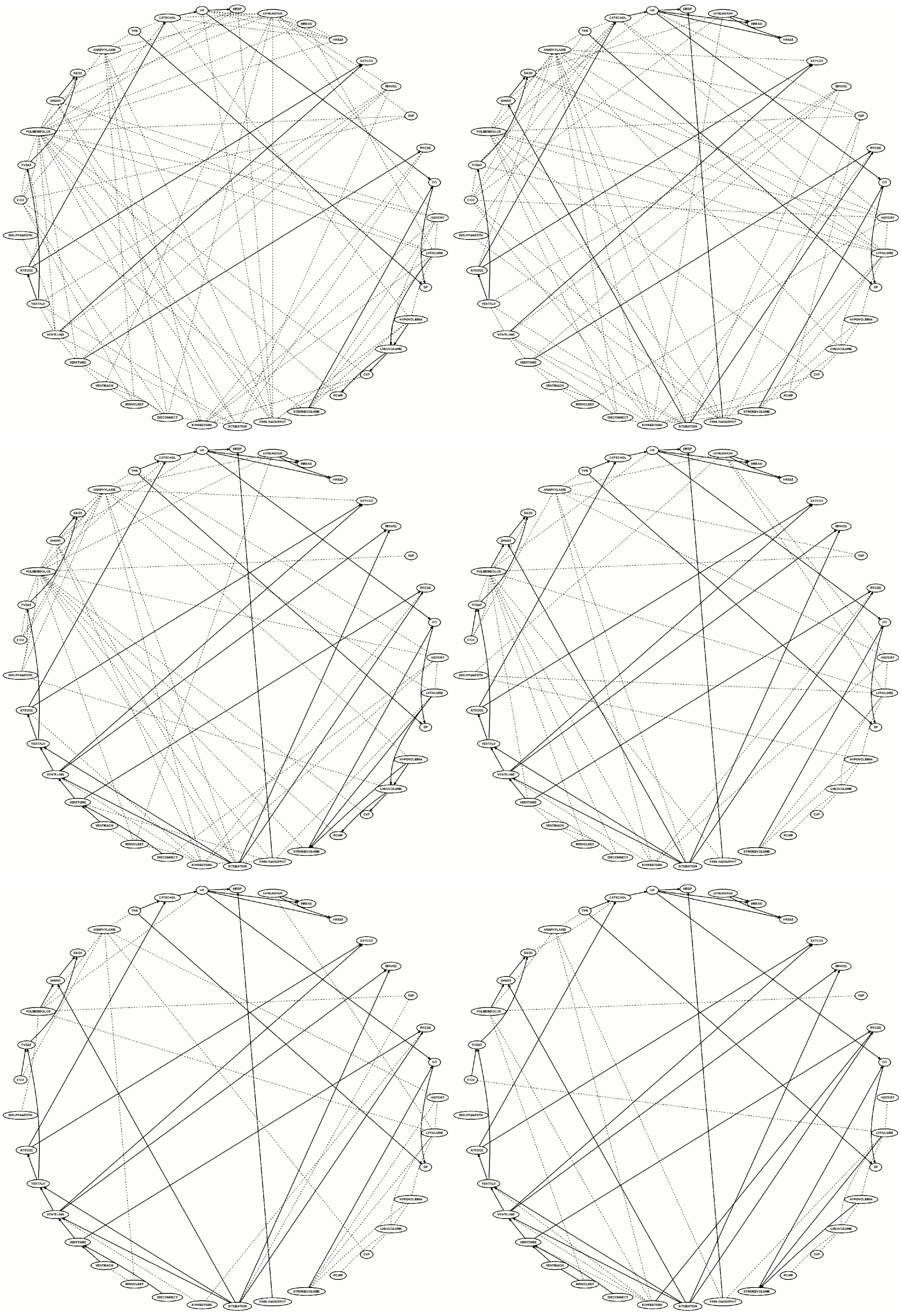

Sachs
Insurance
Alarm
PHD to AG at N = 6400
PHD to AG at N = 12800
PHD to AG at N = 25600
Sample size
BDeu(0.01)
BDeu(0.1)
BDeu(1.0)
BDeu(10)
BIC
fNML

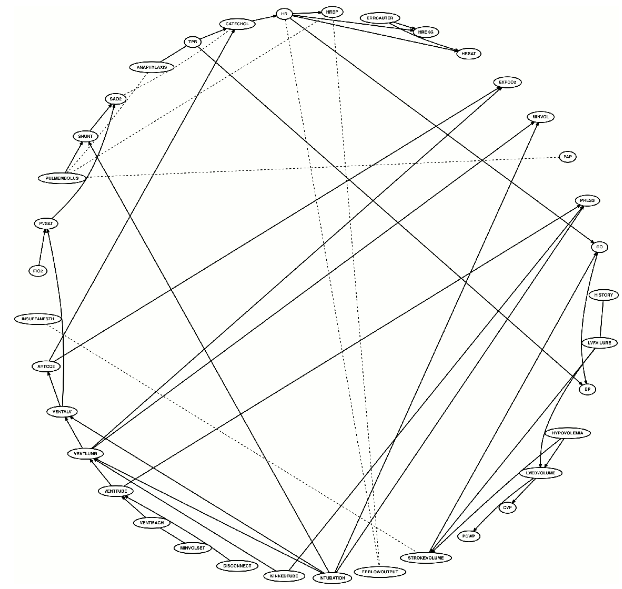
2 Ground truth DAGs, CPDAG, and agreement graphs

Order in all cases from left to right (and top to bottom): DAG, CPDAG, Agreement graph at intersection points 100, 200, 400, 800, 1600, 3200, 6400, 12800, and 25600. All agreement graphs are from the first data subsample.

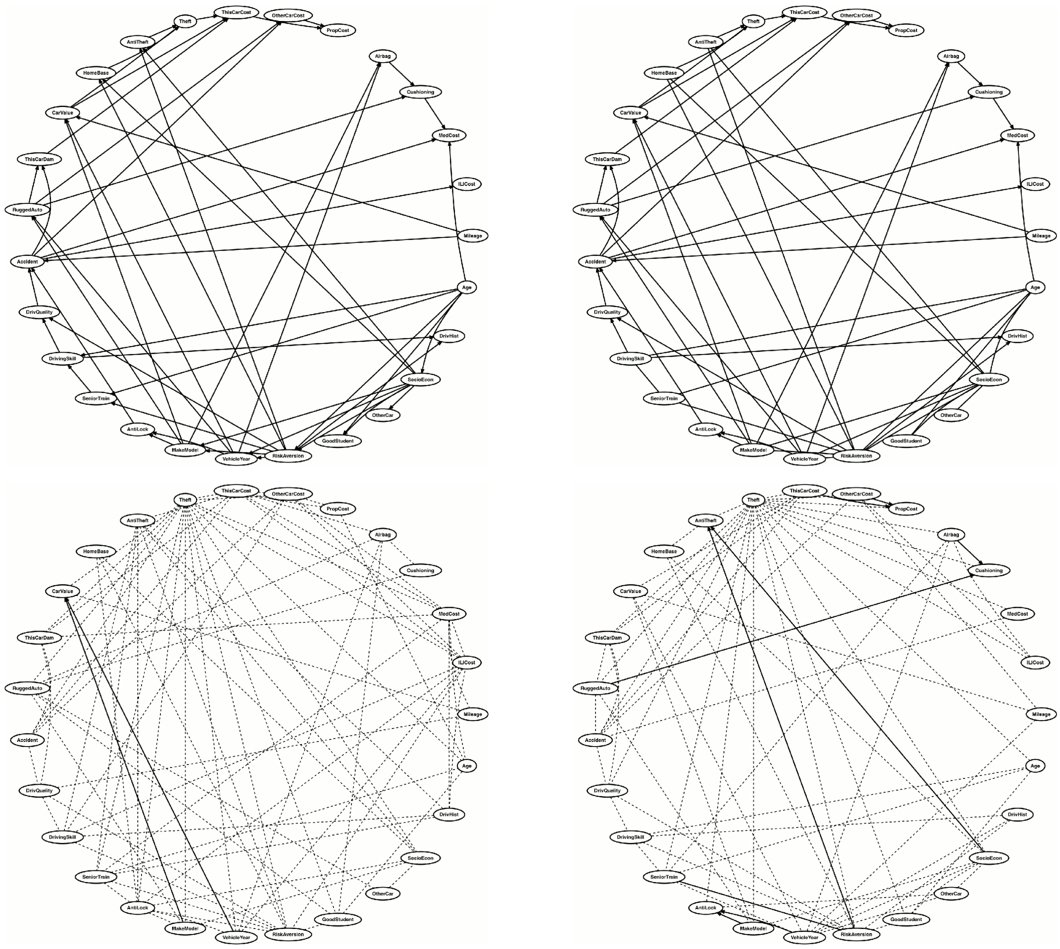
2.1 Alarm

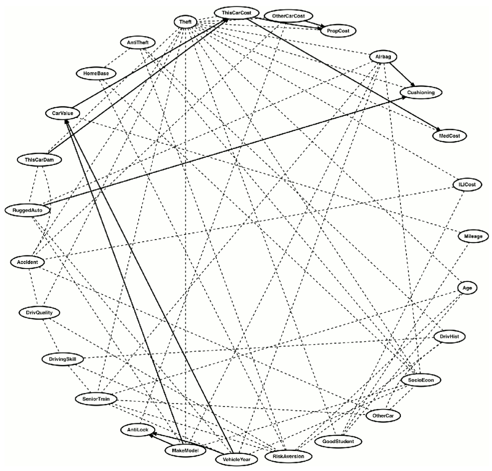


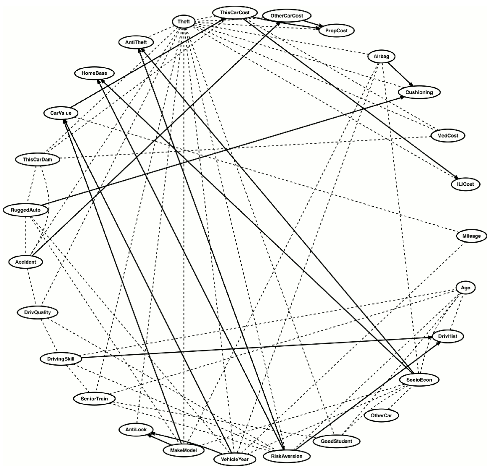

CATECHOL
HR
HRBP
ERRCAUTER
HREKG
HRSAT
TPR
ANAPHYLAXIS
EXPCO2
MINVOL
SAO2
SHUNT
PAP
PULMEMBOLUS
PRESS
PVSAT
CO
FIO2
HISTORY
INSUFFANESTH
LVFAILURE
ARTCO2
BP
VENTALV
HYPOVOLEMIA
VENTLUNG
LVEDVOLUME
VENTTUBE
CVP
VENTMACH
PCWP
MINVOLSET
STROKEVOLUME
DISCONNECT
KINKEDTUBE
INTUBATION
ERRLOWOUTPUT

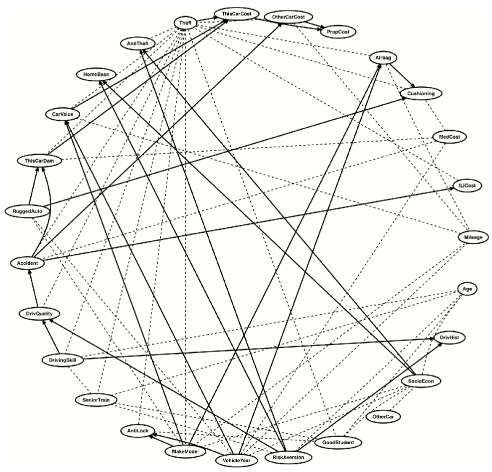


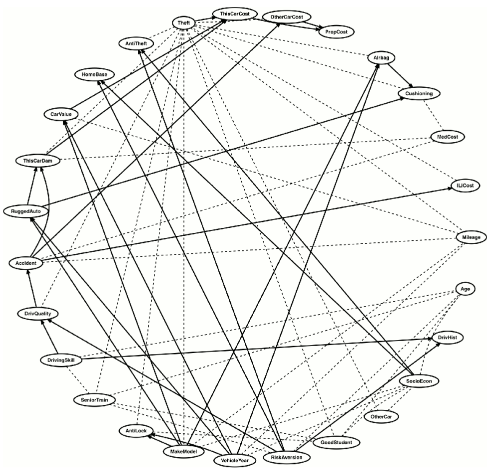
2.2 Insurance

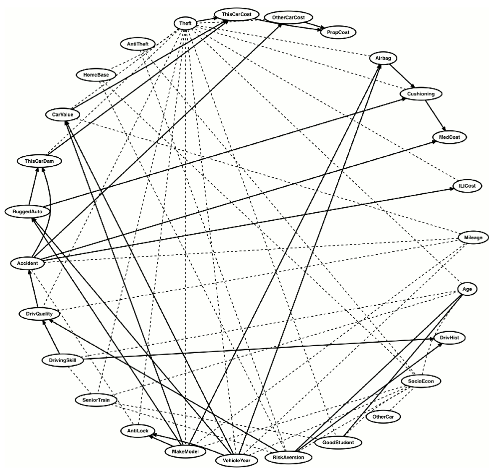


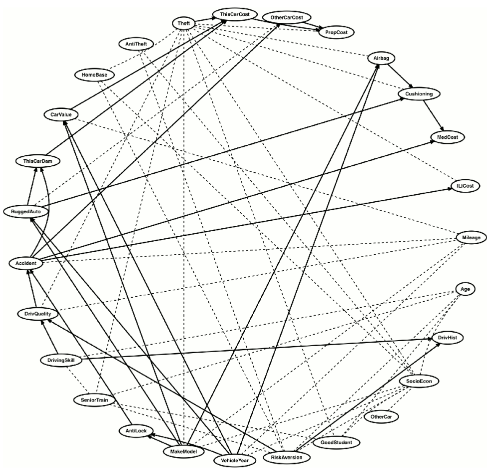

ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft

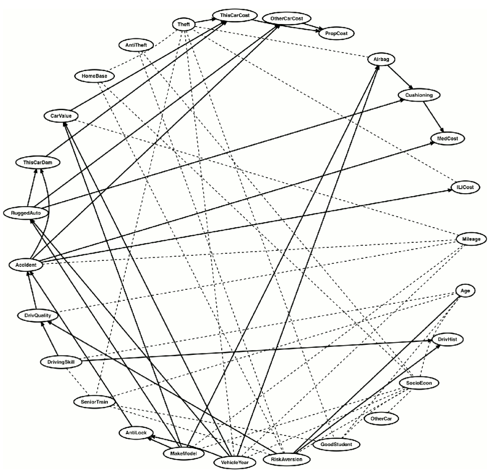

ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft


ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft

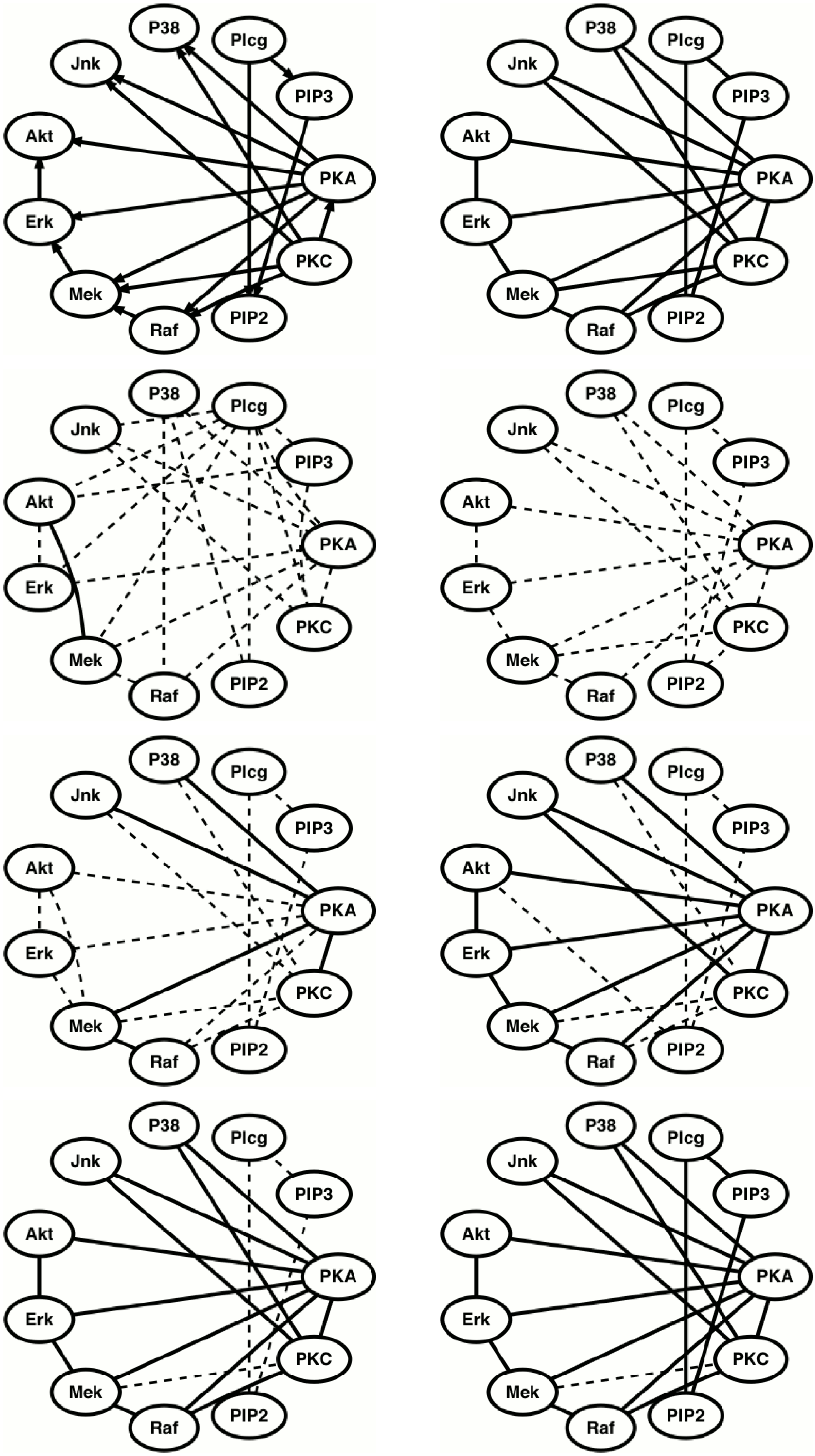

ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft


ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft

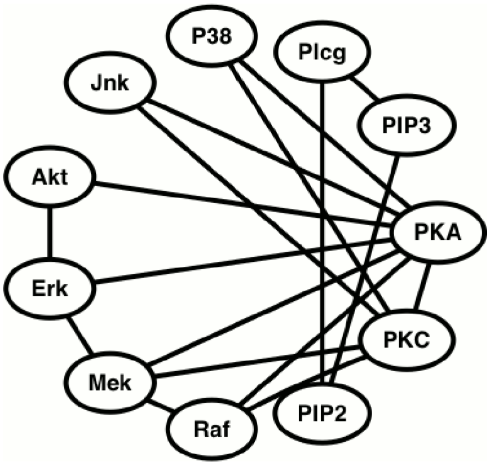

ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft


ThisCarCost
OtherCarCost
PropCost
Airbag
Cushioning
MedCost
ILiCost
Mileage
Age
DrivHist
SocioEcon
OtherCar
GoodStudent
RiskAversion
VehicleYear
MakeModel
AntiLock
SeniorTrain
DrivingSkill
DrivQuality
Accident
RuggedAuto
ThisCarDam
CarValue
HomeBase
AntiTheft
Theft

2.3 Sachs



P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf
P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf


P38
Plcg
Jnk
PIP3
Akt
PKA
Erk
PKC
Mek
PIP2
Raf

3 Performance of intersection-validation in pairwise comparison

To evaluate the performance of intersection-validation in a setting limited to pairs of structure estimators, we ran the procedure (Section 4 of the paper). For each pair the procedure ranks the estimators at different sample sizes using various intersection points for a total of 36 rankings per pair and data set. Contrasting these rankings to the ones obtained by comparing against the ground-truth structure, we calculated the proportion of cases where the two rankings were equal at the corresponding sample sizes. On average, over all estimator pairs and data sets, the proportions is 0.87; for *Sachs*, *Insurance*, and *Alarm*, the mean proportions are 0.87, 0.82 and 0.91 respectively. Tables 1–3 summarize the results for the three data sets and for all pairs of estimators.

Table 1: Accuracy of intersection-validation in pairwise comparison of structure learning algorithms (scoring functions) for *Sachs*

	BDeu ess 0.1	BDeu ess 1	BDeu ess 10	BIC	fNML
BDeu ess 0.01	0.94	0.83	0.83	0.94	0.86
BDeu ess 0.1		0.75	0.83	0.89	0.81
BDeu ess 1			0.97	0.81	0.81
BDeu ess 10				0.86	0.86
BIC					1.00

Table 2: Accuracy of intersection-validation in pairwise comparison of structure learning algorithms (scoring functions) for *Insurance*

	BDeu ess 0.1	BDeu ess 1	BDeu ess 10	BIC	fNML
BDeu ess 0.01	0.92	0.75	0.69	0.92	0.89
BDeu ess 0.1		0.67	0.97	0.94	0.97
BDeu ess 1			1.00	0.89	1.00
BDeu ess 10				0.53	0.28
BIC					0.83

Table 3: Accuracy of intersection-validation in pairwise comparison of structure learning algorithms (scoring functions) for *Alarm*

	BDeu ess 0.1	BDeu ess 1	BDeu ess 10	BIC	fNML
BDeu ess 0.01	0.86	0.97	1.00	0.89	0.89
BDeu ess 0.1		0.83	1.00	0.69	1.00
BDeu ess 1			1.00	0.72	0.97
BDeu ess 10				1.00	1.00
BIC					0.89