

Serafina Di Gioia PostDoc in Applied ML @ ICTP

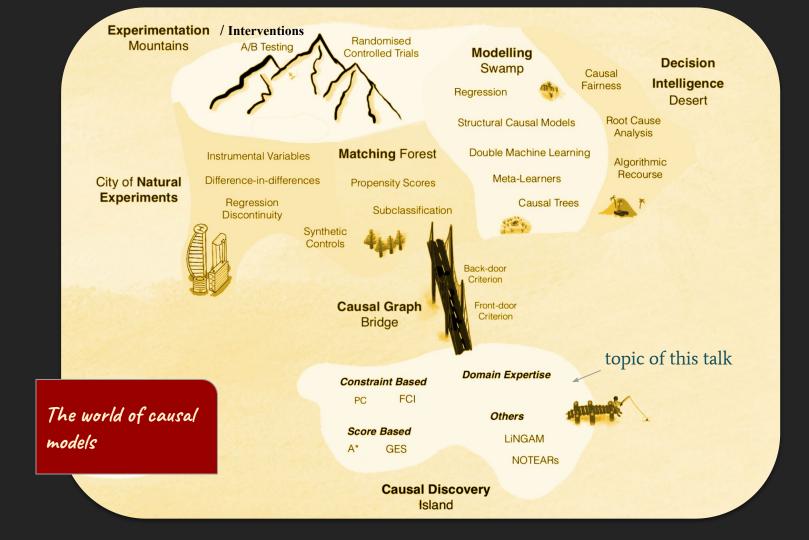
30th November 2023



What is causality? "Actual" Causality "Causality-in-mean"

Statistics

	B	
WA (, J) WA	-	TERFACTUALS
	ACTIVITY:	Imagining, Retrospection, Understanding
	QUESTIONS:	What if I had done? Why? (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)
	EXAMPLES:	Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?
14(2)		
	2. INTER	
	ACTIVITY:	Doing, Intervening
	QUESTIONS:	
	EXAMPLES:	If I take aspirin, will my headache be cured? What if we ban cigarettes?
	1. ASSOC	IATION
	ACTIVITY:	Seeing, Observing
SEEING	QUESTIONS:	What if 1 see? (How are the variables related? How would seeing X change my belief in Y?)
	EXAMPLES:	What does a symptom rell me about a disease? What does a survey rell us about the election results?
	from Boo	ok of Why by Pearl



Tools for causal discovery

• THEORY

Probabilistic graphical models + causal inference

• **COMPUTATIONAL METHODS:** causal structure/parameter learning algorithms

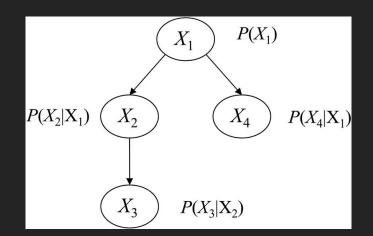
Major ingredient of PGMs: Bayesian networks (BN)

Bayesian networks have 2 components:

• directed acyclic graph (DAG)

• Joint probability distribution

$$\mathbf{P}(\mathbf{X}) = \prod_{i=1}^{N} \mathbf{P}(X_i \mid \Pi_{X_i}; \Theta_{X_i})$$



The world of DAGs

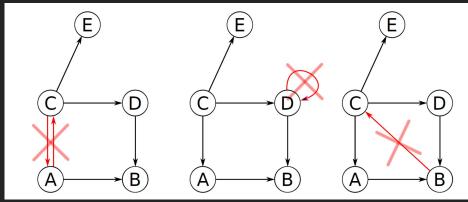
We can indicate a graph with the tuple

$$\mathcal{G} = (\mathbf{V}, A)$$
 set of nodes $a_{ij} = (v_i, v_j)$ set of arcs

If a graph satisfies:

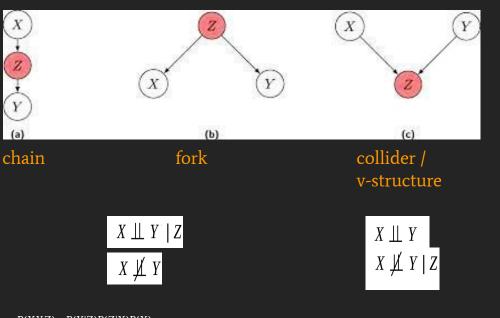
- 1. no undirected edges
- 2. no loop
- 3. no cycle

is called a Directed Acyclic graph (DAG)



Building blocks of causal BNs

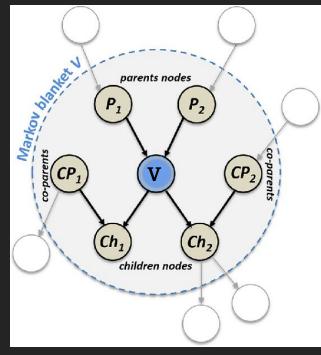
Triplets



P(X,Y,Z) = P(Y|Z)P(Z|X)P(X)=P(Y|Z) P(X|Z)P(Z)

P(X,Y,Z) = P(Z|X,Y) P(X) P(Y)

Markov blanket



Conditional independence is encoded in graph properties

Probabilistic Graphical DAG separation independence $A \perp\!\!\!\perp_P B$ $A \perp\!\!\!\perp_G B$ B $A \perp _P D \mid C$ $A \perp\!\!\!\perp_G D \mid C$ $B \perp\!\!\!\perp_P D \mid C$ $B \perp\!\!\!\perp_G D \mid C$ $A \perp _P E \mid C$ \implies $A \perp\!\!\!\perp_G E \mid C$ E $B \perp\!\!\!\perp_P E \mid C$ $B \perp\!\!\!\perp_G E \mid C$ D $D \perp\!\!\!\perp_P E \mid C$ $D \perp\!\!\!\perp_G E \mid C$ $C \perp\!\!\!\perp_P F \mid D$ $C \perp\!\!\!\perp_G F \mid D$ F

 $\mathbf{A} \perp _{P} \mathbf{B} \mid \mathbf{C} \Longleftarrow \mathbf{A} \perp _{G} \mathbf{B} \mid \mathbf{C}$

Causal Structure-learning algorithms:

There are three main classes of algorithms for causal discovery:

- constraint-based
- score-based
- hybrid

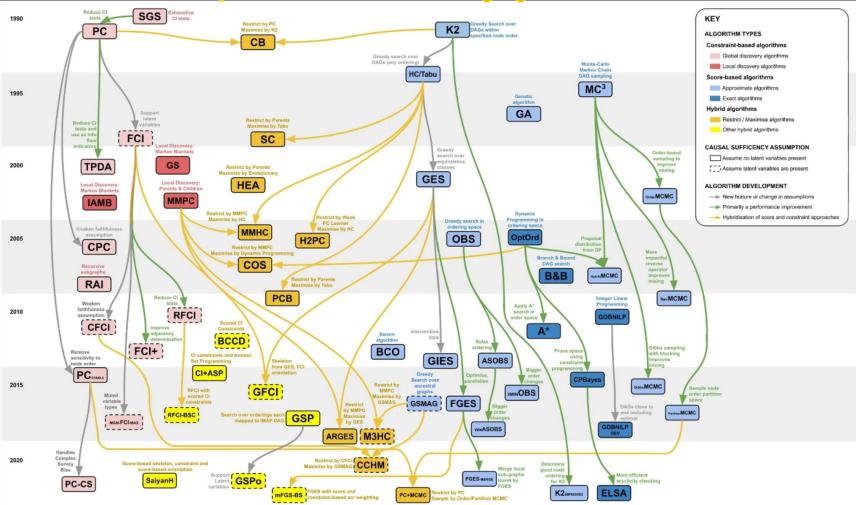
- **Conditional Independence tests**
- → score optimization + grid-search
- ----- mixing score optim. with CI tests

General assumptions:

- 1. Causal Markov
- 2. Causal sufficiency
- 3. Causal faithfulness

no collinear variables no missing confounders

History of causal structure learning algorithms



The gABi library

main - P 2 branches 💿 0	Go to file	Add file - Code -	About
Sera91 Update README.md	883702	ta 1 hour ago 🌀 92 commits	Python-C package for performing bayesian automated inference on da based on probabilistic graphical
gABIC	Update README.md	7 months ago	models.
profiling	new file: profiling/image_zoom.html	last year	Readme
.gitignore	new file: .gitignore	last year	Activity
README.md	Update README.md	1 hour ago	☆ 0 stars
SETUP.md	Update SETUP.md	7 months ago	3 watching 0 forks
gA8I.yml	new file: gABI.yml	8 months ago	¥ utorks
logo-GABI.png	new logo	1 hour ago	Releases
logo-pacchetto-github.png	Add files via upload	last year	No releases published
test_Asia.py	modified: coreBN/build/lib/coreBN/estimators/PC.py	7 months ago	Create a new release
README.md		-	Packages
	gABI		No packages published Publish your first package
			Python \$2.5% Other 0.1% Suggested Workflows
gABi: graphica	l Automated Bayesian infe	erence	Suggested Workflows Based on your tech stack
5 5 1			Suggested Workflows Based on your tech stack Python package Configure Create and test a Python package on
5 5 1	I Automated Bayesian infe		Suggested Workflows Based on your tech stack
Python package to perform Ba	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Python package Configure Create and test a Python package on multiple Python version.
Python package to perform Ba This package has the following • decorator • itertools (built-in)	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Python package Configure Create and test a Python package on multiple Python version.
Python package to perform Ba This package has the following • decorator • itertools (built-in) • matplotlib (built-in	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Python package Contigue Cruits and rest a Python package on multiple Python versions. dj Django Configure
Python package to perform Ba This package has the following • decorator • itertools (built-in) • matplotlib (built-in • networkx (version > 2.6)	I Automated Bayesian infe		Suggestel Workflows Based on your that data Python package Configure Order and lets J Python package on multiple Hython version. Build and Test a Djorgo Project
Python package to perform Ba This package has the following • decorator • itertools (built-in) • matpilotibi (built-in • networkx (version > 2.6) • joblib	I Automated Bayesian infe		Suggesteller Stack Suggesteller Stack Stac
Python package to perform Ba This package has the following • decorator • itertools (built-in) • matplotlib (built-in • networkx (version > 2.6)	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Python package Configure onutging rython versions. di Diang Configure Build and tect a bying Pripet: Dial and tect a bying Pripet: Create and test a Python age/custon.
Python package to perform Ba This package has the following • decorator • Itertools (built-in) • matplotlib (built-in • networks (version > 2.6) • joblib • os	I Automated Bayesian infe		Suggested Montflows Based on your tech stack Posten and test a Aytom package Configure multiple Aytom services. dj Django Configure Build and Tetta Django Project Ø Phono application Configure Oreste and test a Aytom spelication.
Python package to perform Ba This package has the following decorator iterotols (built-in) matplotib (built-in) networks (version > 2.6) jobilib os pandas	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following • decorator • Iterools (built-in) • matpiobilib (built-in) • networkx (version > 2.6) • jobilb • os • jandas • pypickle	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Python package Configure onutging rython versions. di Diang Configure Build and tect a bying Pripet: Dial and tect a bying Pripet: Create and test a Python age/custon.
Python package to perform Ba This package has the following e decorator iterools (built-in) matpioliti (built-in) e networkx (version > 2.6) o 3 e pandas e pypickile e scipy e statsmodels t ddm	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following decorator itercols (built-in) matpiolitik (built-in) networko (version > 2.6) jobilb os pandas ppytokle scipy statsmodels tadm wget	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following e decorator intertools (built-in) matapiotitis (built-in) matapiotitis (built-in) matapiotitis (built-in) matapiotitis (built-in) e so papital so pypital- statsmodels statsmodels pyparow pyparow	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following decorator intercoles (built-in) matpiolitik (built-in) networko (version > 2.6) jobilb os pandas piprickle scipy scipy statismodels tidm wyget parrow Apache arrow	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following • decorator • iterools (bullt-in) • matpiolitik (bullt-in) • networks (version > 2.6) • jobilb • os • pandas • pypickle • statsmodels • statsmodels • statamodels • statamode	I Automated Bayesian infe		Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange
Python package to perform Ba This package has the following decorator intercoles (built-in) matpiolitik (built-in) networko (version > 2.6) jobilb os pandas piprickle scipy scipy statismodels tidm wyget parrow Apache arrow	I Automated Bayesian infe		Suggested Workflows Based on your tesh stack Python package Configure Create and rest. 3 Python package on Create and rest. 3 Python package on Build and test a Spinger Project Distance of the stack of the stack Python application Configure Create and rest a Spinon application
Python package to perform Ba This package has the following decorator intercols (built-in) matpiolitik (built-in) networks (version > 2.6) joblib os pandas pypickle scipy scipy statsmodels todm paravel yaparovi Apache arrow pybind11 dask intercols and a statsmodels intercols arrow pybind11 dask intercols and a statsmodels intercols arrow intercols arrow intercols and a statsmodels intercols arrow pybind11 dask intercols arrow intercols arrow inte	I Automated Bayesian infe	graphical models.	Suggested Workflows Based on your tesh stack Python package Configure Create and rest. 3 Python package on Create and rest. 3 Python package on Build and test a Spinger Project Distance of the stack of the stack Python application Configure Create and rest a Spinon application
Python package to perform Ba This package has the following decorator intercols (built-in) matpiolitik (built-in) networks (version > 2.6) joblib os pandas pypickle scipy scipy statsmodels todm paravel yaparovi Apache arrow pybind11 dask intercols and a statsmodels intercols arrow pybind11 dask intercols and a statsmodels intercols arrow intercols arrow intercols and a statsmodels intercols arrow pybind11 dask intercols arrow intercols arrow inte	I Automated Bayesian infe yesian causal discovery on data, based on probabilistic dependencies:	graphical models.	Suggested Workflows Based on your tech stack Tython package Configure order and test a Sylma package on multiple rython serions. di Diango Configure Build and test a Dynap Project Diange Configure Diange



The gABI library provides three constraint-based algorithms for causal structure learning:

PC

- kernel-PC
- IAMB

Comparison between PC-stable and kernel-PC

PC-stable is the most common constraint-based algorithm and kernel-PC is its generalization to non-linear ANMs

	Kernel-PC	PC-stable
Differences in assumptions	non-linear relationship gaussian/non-gaussian noise	linear relationships gaussian noise gaussian data
differences in implementation	generalized transitive phase	transitive phase
Conditional Independence test	HSIC-gamma /DCC-gamma HSIC-perm / DCC-perm	Pearson CI test Fisher-Z

Digression on CI tests

For LINEAR GAUSSIAN models described by

$$X_i = \beta_j X_j + \epsilon_i$$

we use PARAMETRIC CI tests (e.g. Fisher-Z test, Exact-t test)

For NON-LINEAR ADDITIVE NOISE models described by

$$X_j = f_j(\mathrm{Pa}_j) + N_j$$

we use kernel NON-PARAMETRIC tests (e.g. HSIC-perm, HSIC-gamma, DCC-gamma)

but in practice...

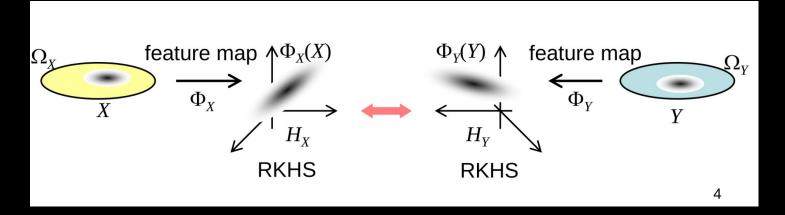
- estimating conditional dependence for continuous domain is not straightforward
- many algorithms use linear statistical methods or discretization

Why kernel methods for testing CI?

basic idea

covariance structure on RKHS gives info on dependence and conditional independence of original variables $\langle g, \Sigma_{YX} f \rangle_{\mathcal{H}_{\mathcal{Y}}} = \operatorname{Cov}[f(X), g(Y)]$

can be decomposed $\Sigma_{YX} = \Sigma_{YY}^{1/2} V_{YX} \Sigma_{XX}^{1/2}$ $V_{YX|Z} = V_{YX} - V_{YZ} V_{ZX}$ $V_{YX|Z} = \Sigma_{YY}^{-1/2} (\Sigma_{YX} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZX}) \Sigma_{XX}^{-1/2}$



Kernel CI tests: residuals gamma test

Test hypothesis:

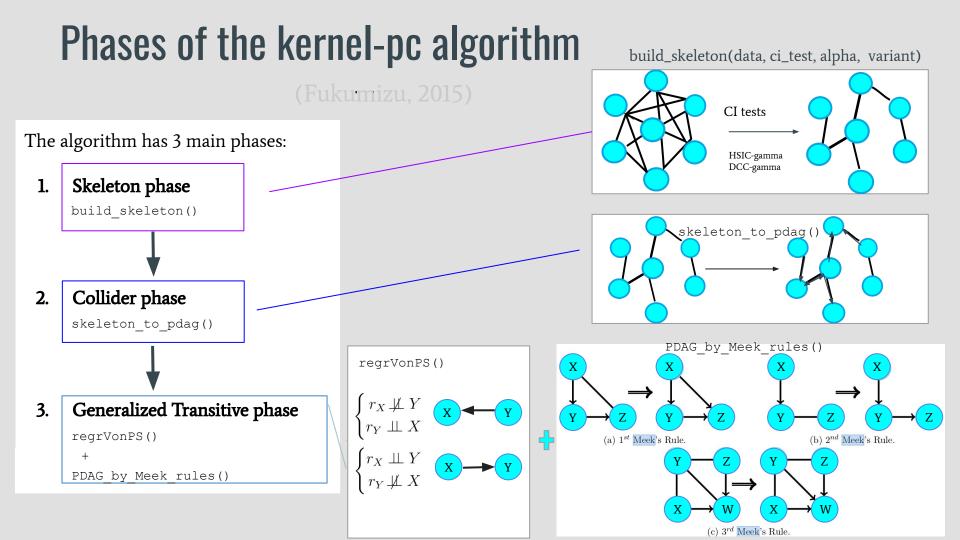
 $\mathcal{H}_0: X \perp\!\!\!\perp Y | Z$ **v.s.** $\mathcal{H}_1: X \not\!\!\!\perp Y | Z$



residuals are estimated with a GAM fit $y = f_0 + \sum_{i=1}^p f_i(x_i) + \varepsilon$ (pyGAM)

Gamma approximation

 $IC(x, y) \sim Gam(\alpha, \theta) \quad \text{where} \quad \alpha = \frac{E[\hat{IC}_{X,Y}]^2}{Var(\hat{IC}_{X,Y})}, \quad \theta = \frac{Var(\hat{IC}_{X,Y})}{E[\hat{IC}_{X,Y}]}$ p-value: upper-tail quantile of Gam $\mathbb{H}(P, Q) = \frac{1}{(N-1)^2} \text{tr}(K_P H K_Q H) \quad \text{where} \quad K_{i,j} = \exp \frac{(x_i - x_j)^2}{\sigma^2}, \quad H_{i,j} = \delta_{i,j} - \frac{1}{n}$



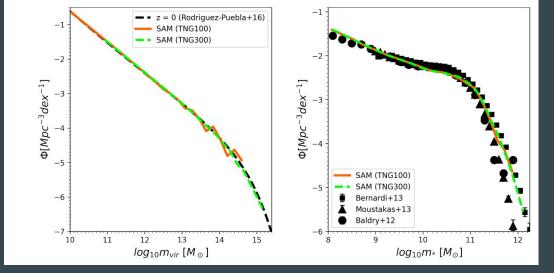
First application: what causes disk size evolution ?

simulated data from

SAM : Santa-Cruz TNG

(Gabrielpillai, 2021)

Collaborators



- V. Acquaviva (Prof @ CUNY)
- F. Bucinca (PhD graduate, CUNY)
- A. Maller (CUNY/CCA)
- R. Trotta (SISSA)

Pipeline of the analysis

- 1. Simulated galaxy catalog creation
- 2. preprocessing
- 3. subsampling
- 4. run kernel-PC on every subsample
- 5. estimate average reconstructed graph with 10 bootstrap
- 6. apply pySR to derive quantitative laws/relations

Preprocessing

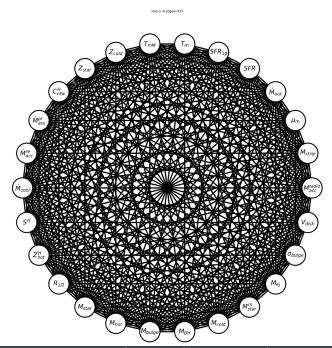
- 1. subsetting data by galaxy morphology (disks / ellipticals) $M_{bulge}/M_{star} < 0.4$
- 2. identification of collinear variables

GalpropMaccdot_radio -	1	0													0	0.9999																					0
GalpropMstrip -	0	1	0																																		0
GalpropMu_merger -	0		1	0																							٥	0									0
GalpropOutflowRate_Mass	0		0	1	٥	0																					0 0	.9617									0
GalpropOutflowRate_Metal	0				1	0.9545	•	0	0																												0
GalpropSfr -	0				0.9545	1	0.9652	0	0.9899																												0
GalpropSfrave100myr -	0					0.9652	1	0	0.982																												0
GalpropSfrave1gyr -	0					0	0	1	0																												0
GalpropSfrave20myr -	0					0.9899		0	1	0																											0
GalpropTmerger -	0								0	1	0																										0
GalpropTmerger_major	0									0	1	0																									0
GalpropZcold -											0	1	0					•	0																		0
GalpropZstar -	0											0	1	0					0.967																		0
HalopropC_nfw -													0	1	0																						0
HalopropMaccdot_pristine -	0													0	1	0																					0
HalopropMaccdot_radio -	-			0	0			0				0			0	1	0		0	0			0						0	0			0	0			0
HalopropMaccdot_reaccreate	0			0	•				•			0				0	1	0	0	•									•				•	•			0
HalopropMcooldot -													0				0	1	1																		0
HalopropMetal_ejected	0											0	0.967					0	0																		0
HalopropSpin - HalopropZhot -	0																		0	1	0		0														0
GalpropNormHalfRadius -																			0	0	0	1															0
GalpropNormMstar	. 0																				0	0	1			.9882											0
HalopropNormMhot -	0																						0	1	0	0											0
GalpropNormMbulge -	0																						0	0	1												0
GalpropNormMdisk -	0																					0 0	0.9882	0	0	1											0
GalpropNormMBH -	0																					0	o		0	0	1										0
GalpropNormMH2	0		0	0.9617	0																					0	0	1									0
GalpropNormMHI -	0			0																							0	0	1 0	.9556 (0.9535						0
GalpropNormMHII -	0																											0 0	.9556	1	0						0
GalpropNormMcold -	0																											0 0	.9535	0	1						0
GalpropNormMstar_merge -	0																											٥	0	0	0	1	o	0.9758			0
HalopropNormMass_ejected	0																														0	0	1	0			0
HalopropNormMstar_diffuse -	0																														0	0.9758	0	1	0		0
GalpropNormSigmaBulge -	0																																	0	1	0	0
GalpropNormVdisk	0																																		0	1	0
fdisk -	-	0	0	0	0	٥	0	٥	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	٥	0	0	0	٥	0	٥	٥	0	0		1
	radio	Mstrip	nerger	Mass	Metal	GalpropSfr	00myr	elgyr	20myr	nerger	major	oZcold	pZstar	C_nfw	ristine	radio	create	coldot	ejected	pSpin	pZhot	Radius	Mstar	nMhot	Mbulge	Mdisk	MBH	mMH2	IHWI	IIHWW	Mcold	merge	ejected	diffuse	sBulge	nvdisk	fdisk
	Maccdo	GalpropMstrip	opMu_merge	lowRate	owRate	Galp	Sfrave1	GalpropSfrave1gy	GalpropSfrave20my	GalpropTmer	merger	GalpropZcold	GalpropZsta	HalopropC,	ppMaccdot_prist	opMaccdot	cdot_reaccre	propMc	opMetal_e	Halopre	HalopropZh	ormHalt	GalpropNormMstar	HalopropNormMhol	E	GalpropNormMdisk	GalpropNormMBH	GalpropNormMH2	IpropNo	GalpropNorr	GalpropNormMcold	opNormMstar_merge	Mass_e	Mstar_diff	msigm	GalpropNormVdisk	
	GalpropMaccdot_radio	-	Galpri	GalpropOutflowRate_Mass	GalpropOutflowRate_Meta		GalpropSfrave100my	Galpr	Galproj	5	GalpropTmerger_majo			Т	propMac	HalopropM	opMaccde	Halo	dd			GalpropNormHalfRad	Galpr	Halop	GalpropNor	Galpr	Galp	Galp	Galp	Galg	Galpr	opNorm	opNorm	phorm	GalpropNormSigmaBulge	Galpa	
	0			Galpr	Galpro		0				3				Halop	1	taloprop		Hal			3										Galpri	Halopro	Halopro	Galp		
																	r																				

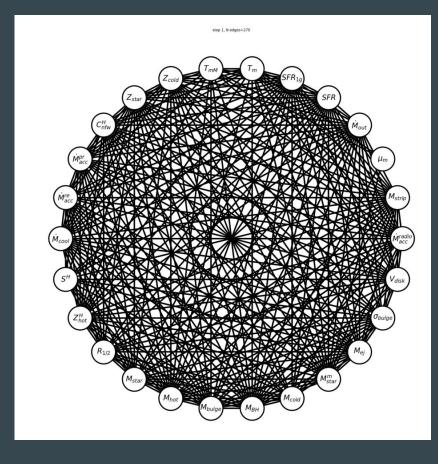
After preprocessing -> we are left with 26 vars out of 37

Name	description	Name	description
M_{strip}	stripping mass	SFR_{1g}	$\langle SFR \rangle_{1Gyr}$
$M_{outflow}$	outflow rate	T_{merge}	T last merger
μ_{merger}	mass ratio mergers	T_{mergeM}	T last major merger
Z_{cold}	metallicity cold gas	$\dot{M_{acc}^{pr}}$	pristine accretion rate
$\dot{M_{acc}^{re}}$	reaccretion rate	C_{nfw}	halo concentration
$Spin^{H}$	halo spin	$\dot{M_{cool}}$	cooling mass
$R_{1/2}$	half-mass radius	Z_{hot}^H	metallicity hot gas
M_{BH}	BH mass	M_{cold}	CG mass
$M_{ejected}$	Ejected Mass	V_{disk}	disk vel.
M_{star}	stellar mass	M_{hot}	Hot gas mass
σ_{bulge}	bulge vel. disp.	SFR	Star formation rate
M_{bulge}	bulge mass	Z_{star}	stellar metallicity
M_{star}^{merge}	stellar mass from merg.	\dot{M}_{acc}^{radio}	radio accretion rate

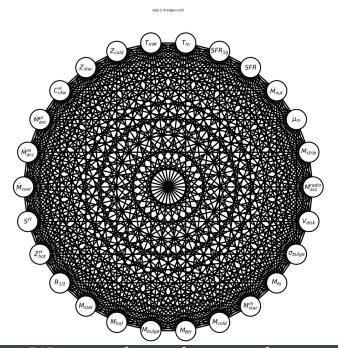
Skeleton building phase



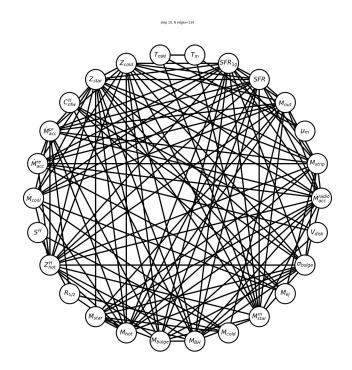
We start from the moral graph and we stop at S_set size=10



Skeleton building phase

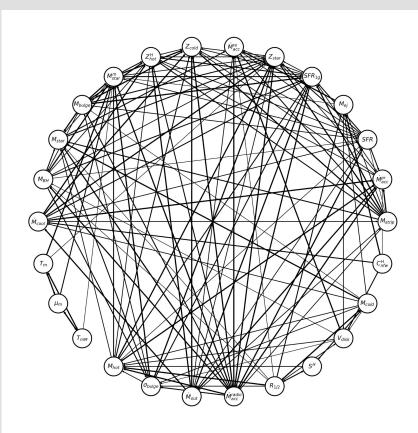


We start from the moral graph and we stop at S_set size=10



Average reconstructed BN

We adopt the approach proposed by <u>Friedman, Goldszmidt and Wyner (1999):</u> generating multiple network structures by applying nonparametric bootstrap to the data and estimating the relative frequency of the feature of interest



BN averaged over 10 bootstrap subsamples (N=4000)

Future research directions

After concluding the current study on causal graphical tools applied to galaxy simulated data we plan to move to real data, and we plan to integrate in the gABi package:

- treatment for latent variables (based on SEM packages, e.g. SEMOPY)
- treatment for time-series data (already under implementation)

For more general application we will add to gABi package:

- treatment for mixed categorical/continuous variables
- new CI tests: Shrinkage test
- new possible moves for the CI sampling, like the so-called new edge reversal move (Grzegorczyk and Husmeier, 2008)
 - and Husmeler, 2008)
- new structure-learning algorithms: layering-MCMC*, hill-climbing, greedy search

This will allow to perform Bayesian inference on time-series data, even in the case where latent confounders are present.

LONG-TERM GOAL: causal model discovery on real data

References

Edwards DI (2000). Introduction to Graphical Modelling. Springer, 2nd edition

Fukumizu(2007) Kernel Measures of Conditional Dependence

Tsamardinos I, Borboudakis G (2010). "Permutation Testing Improves Bayesian Network Learning". Machine Learning and Knowledge Discovery in Databases, 322–337

Tillman, R. E., Gretton, A. and Spirtes, P. (2009). Nonlinear directed acyclic structure learning with weakly additive noise model. NIPS 22, Vancouver

Colombo D, Maathuis MH (2014). "Order-Independent Constraint-Based Causal Structure Learning". Journal of Machine Learning Research, 15:3921–3962. (article introducing pc-stable)

Matthew J. Wovels et al. (2021), D'ya like DAGs? A Survey on Structure Learning and Causal Discovery, arXiV 2103.02582v2

Russell SJ, Norvig P (2009). Artificial Intelligence: A Modern Approach. Prentice Hall, 3rd edition



The End