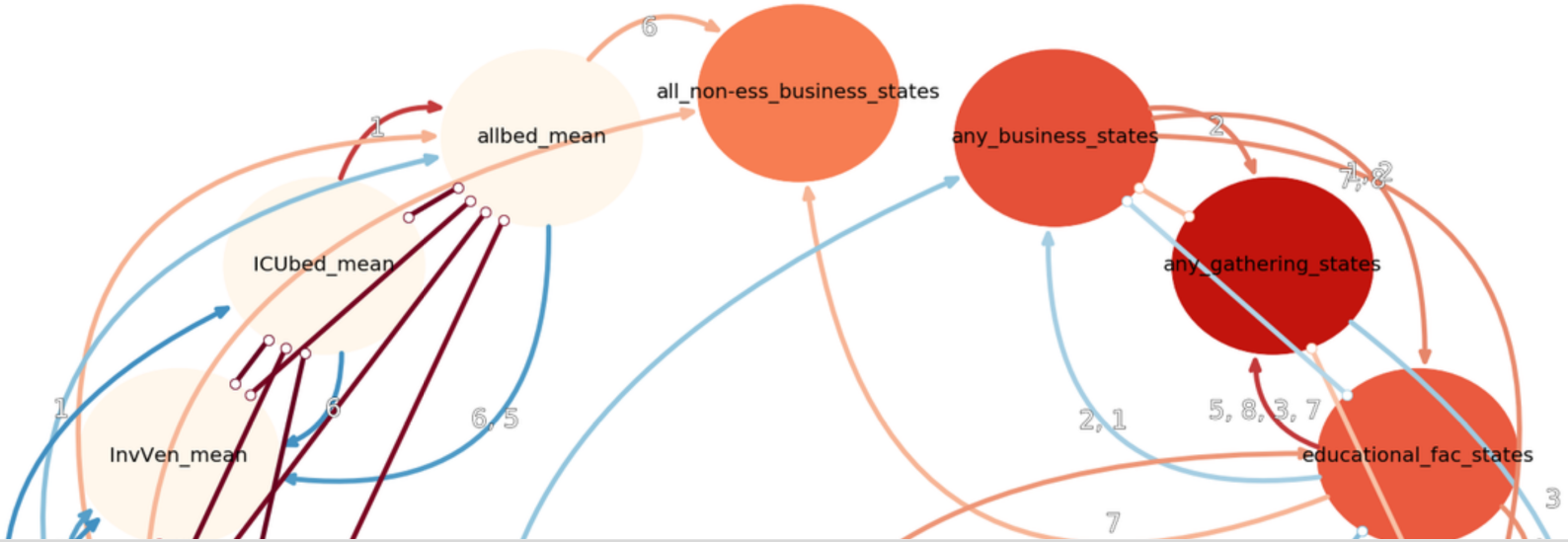
Karlsruher Institut für Technologie

Smart Data Analytics
Assignment 4 – Causal Inference

Zihan Chen, Thassilo Helmold, Gerrit Merz, Robin Schnaidt – Team 4



Content

Motivation

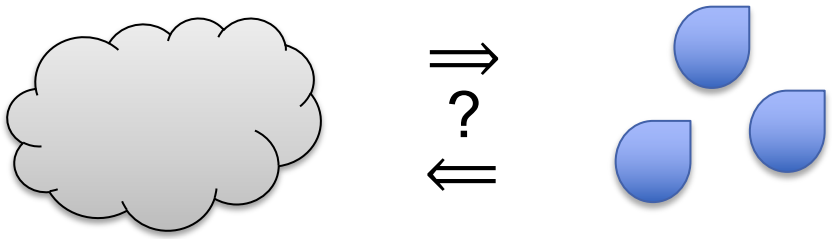
- Obtain knowledge through experiments

https://home.cern/sites/home.web.cern.ch/files/2018-10/lhc-tunnel-2302977.jpg

- But: *What to do if experiments are not feasible?*

Causal Inference

- **“Correlation does not imply causation”**
- Weather Example: High correlation between rain and number of clouds



- **…but which phenomenon causes which one?**
- Correlation is *symmetric* measure: no sense of direction

- Causality: "A signal X is said to cause Y if the **future realizations of Y** can be better explained using the **past information from X and Y** rather than Y alone" (Granger 1969)

- Need for *asymmetric* measures

Transfer Entropy

- R Package "RTransferEntropy"
- Shannon entropy
- Rényi entropy

PCMCI (Runge 2019)

- Python Package "Tigramite"
- Iterative conditional independence testing
- Can be used with different Independence tests

Temporal Causal Discovery Method (TCDF)

- Uses Attention-based Convolutional Network

Causeme

- Evaluation with the Causeme-Benchmark-Platform
- No Ground-Truth -> Stick to given metrics

- Single benchmark can contain up to 200 datasets
- One benchmark group contains many benchmarks of different complexity

- => Need to conduct *many* experiments

SDIL-Batch-System

Smart Data Innovation Lab

Karlsruher Institut für Technologie

- Use SDIL-Batch-System to efficiently execute many experiments in parallel
- Use modified sample script (command line parameters)
- Progress reporting via temporary file names

- Single experiments still take a long time, *up to 12 hours and more*

```
pcmci_linear-VAR_aggregated_N-40_T-150_agg-5_0159.txt
pcmci_linear-VAR_aggregated_N-40_T-300_agg-2_0152.txt
pcmci_linear-VAR_aggregated_N-40_T-300_agg-3_0121.txt
tcdf_linear-VAR_aggregated_N-20_T-150_agg-2_0154.txt
tcdf_linear-VAR_aggregated_N-20_T-150_agg-3_0123.txt
```

Evaluation

- Which benchmark datasets to use?
- Which properties does our COVID-dataset have?

- Algorithm should work well for all datasets (robustness)

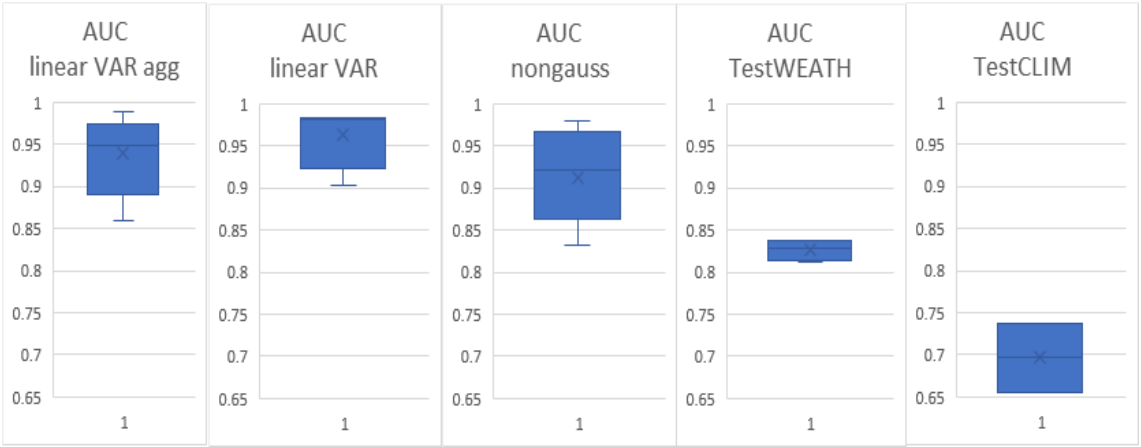
- Evaluation of metrics on small datasets from very different benchmark groups:
  - Linear-VAR
  - Linear-VAR-aggregated
  - TestWEATH
  - TestCLIM
  - river-runoff

PCMCI

Generally good performance
Very sensitive to length of data

Linear-VAR	AUC	F	FPR	TPR
T-150	0.9038	0.8295	0.0533	0.5983
T-300	0.9824	0.9379	0.055	0.9167

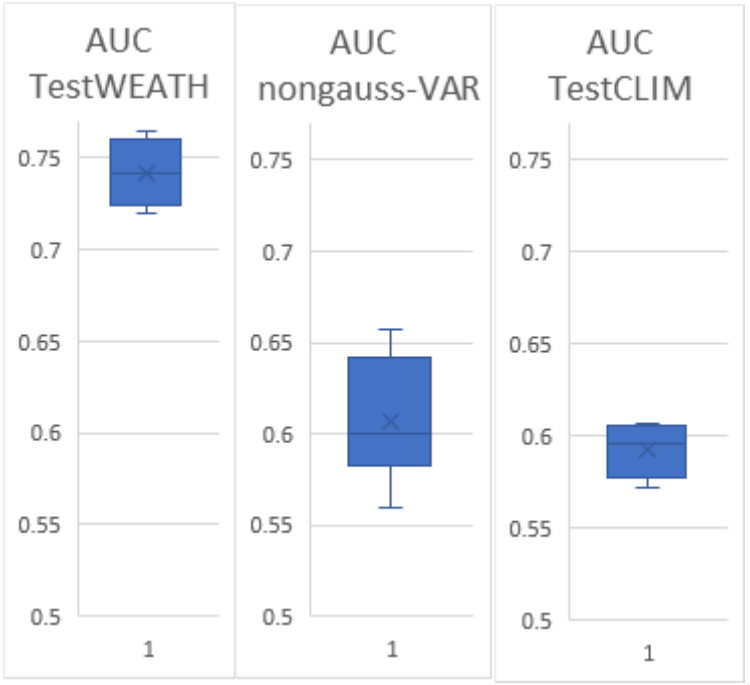
nongauss-VAR	AUC	F	FPR	TPR
T-150	0.8326	0.7986	0.0627	0.508
T-200	0.8968	0.9211	0.0618	0.6672
T-300	0.9622	0.9406	0.0582	0.855



+	-
Fast runtime on small/medium datasets Good Performance on many benchmarks Few hyperparameters	Slow with nonlinear CI-Tests Slow for large numbers of variables

Transfer Entropy

- R Package "RTransferEntropy"
- Runtime on local machine for calculating pairwise transfer entropy (calc_te) for one benchmark ~ 5min



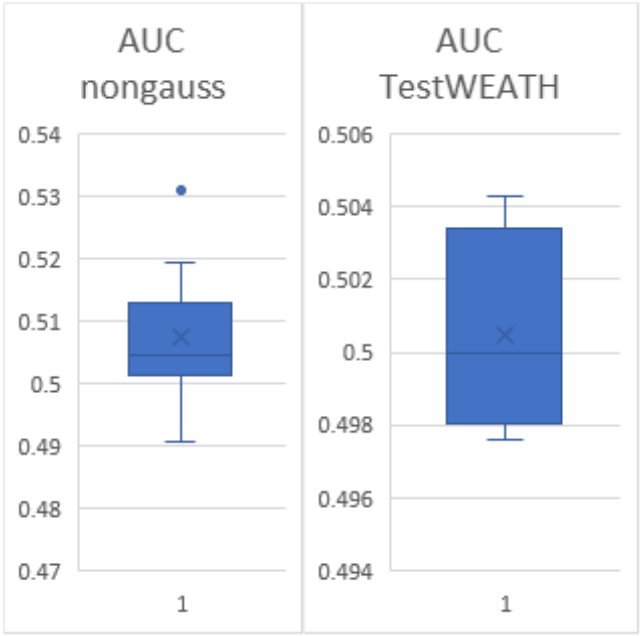
+	-
Very fast runtime when calculating only transfer entropy	No p-values function Overall worse performance than PCMCI

Temporal Causal Discovery Framework (TCDF)

Generally bad / medium performance

Sometimes longer series led to worse performance

Linear-VAR	AUC	F	FPR	TPR
T-150	0.5333	0.4918	0.2717	0.3383
T-300	0.4858	0.4015	0.2783	0.25

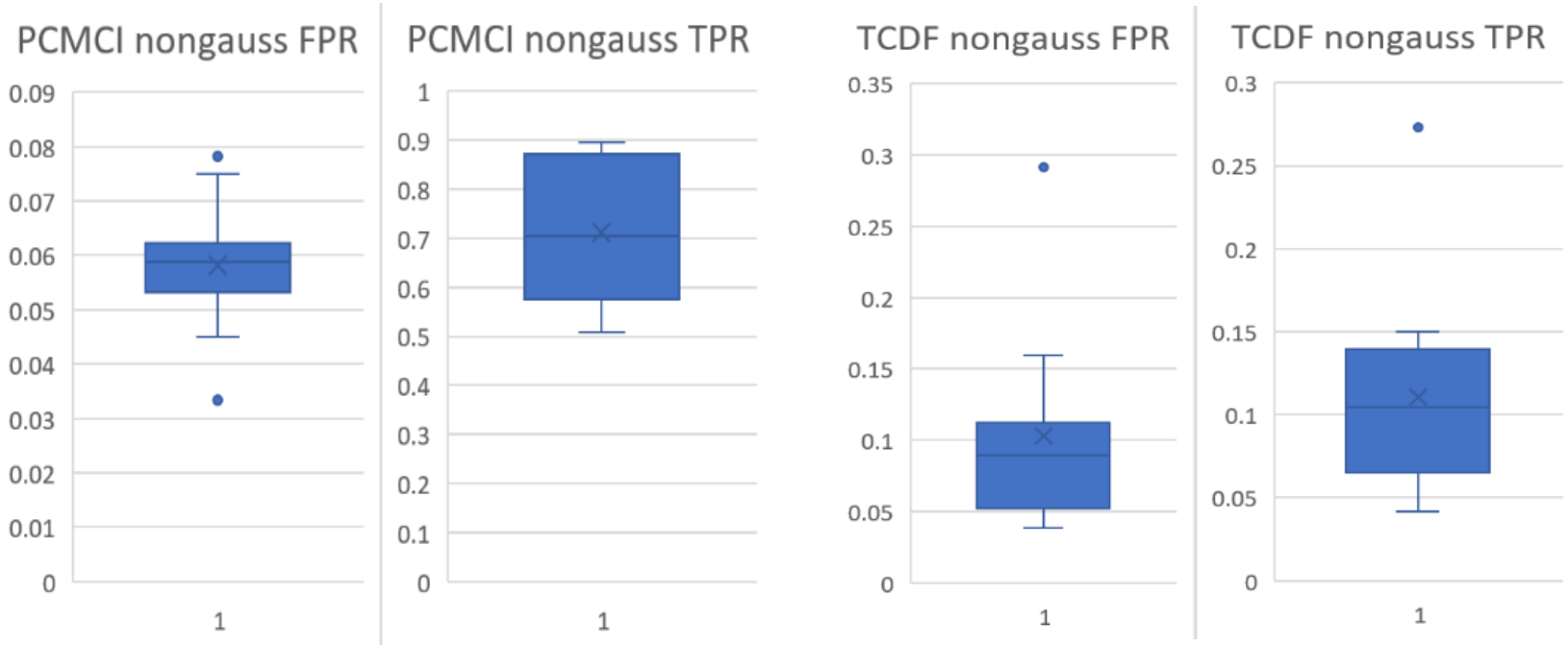


+	-
Medium runtime on small/medium datasets Gives back a set of Causal Links Can tune many hyperparameters	Generally slow Generally bad performance Extremely slow with “better” settings Must tune many hyperparameters

Benchmark-Groups

nongauss-VAR: Linear vector-autoregressive time series model with gaussian and non-gaussian noise

12 Benchmarks, maximum N=20



Benchmark-Groups

TESTWEATH: Nonlinear datasets of weather data

4 benchmarks, maximum N=10

nonlinear -> PCMCI-ParCorr not really suitable (but GPDC, CMIknn too slow)

Experiment	Method	Runtime (s)	AUC	F-measure	FPR	TPR
TestWEATH_N-10_T-1000	test_pcmci	10,2872	0.8201	0.5388	0.0846	0.6047
TestWEATH_N-10_T-1000	test_tcdf	90,9068	0.5006	0.067	0.0203	0.0216
TestWEATH_N-10_T-2000	test_pcmci	12,8951	0.8371	0.5476	0.0886	0.6564
TestWEATH_N-10_T-2000	test_tcdf	62,6077	0.5043	0.0815	0.0166	0.0252
TestWEATH_N-5_T-1000	test_pcmci	3,2136	0.8126	0.6698	0.1052	0.6203
TestWEATH_N-5_T-1000	test_tcdf	26,7976	0.4976	0.1019	0.0355	0.0308
TestWEATH_N-5_T-2000	test_pcmci	3,2019	0.837	0.6684	0.129	0.6971
TestWEATH_N-5_T-2000	test_tcdf	12,9293	0.4994	0.1191	0.0388	0.0376

TCDF Hyperparameters

- TCDF delivers medium/bad results on all our benchmarks
- Maybe the default hyperparameters are just bad for these?

- Kernel size has to fit to the maximum time-lag in the dataset.

- Search over different configurations of key hyperparams:

setting_nr	layers	epochs	lr	significane
0	0	1000	0.01	0.8
1	1	1000	0.01	0.8
2	0	5000	0.01	0.8
3	0	5000	0.001	0.8
4	0	1000	0.01	0.9

TCDF Hyperparameters

Setting	Time	AUC	F-Score	FPR	TPR
setting_0	8,7968	0.5333	0.4918	0.2717	0.3383
setting_1	15,1052	0.5375	0.5076	0.3033	0.3783
setting_2	43,5577	0.5308	0.4935	0.2917	0.3533
setting_3	41,7921	0.5317	0.4816	0.25	0.3133
setting_4	8,1738	0.5358	0.5018	0.2917	0.3633

linear-VAR_N-3_T-150

Setting	Time	AUC	F-Score	FPR	TPR
setting_0	6,3644	0.5192	0.4721	0.2883	0.3267
setting_1	10,2512	0.5017	0.4698	0.37	0.3733
setting_2	31,8925	0.5217	0.473	0.2783	0.3217
setting_3	31,1702	0.5017	0.4457	0.3033	0.3067
setting_4	6,4375	0.5175	0.4913	0.36	0.395

linear-VAR_aggregated_N-3_T-300_agg-3

Setting	Time	AUC	F-Score	FPR	TPR
setting_0	91,102	0.5455	0.3333	0.0	0.0909
setting_1	212,4243	0.6322	0.5556	0.0083	0.2727
setting_2	595,2961	0.6736	0.5714	0.0165	0.3636
setting_3	500,9142	0.5909	0.5263	0.0	0.1818
setting_4	92,5449	0.6322	0.5556	0.0083	0.2727

River-Runoff

No clear results, optimal hyperparams depend on the concrete dataset

But: *An additional hidden layer often leads to more detected causal links*

Causal Discovery for COVID-19 Data Sets

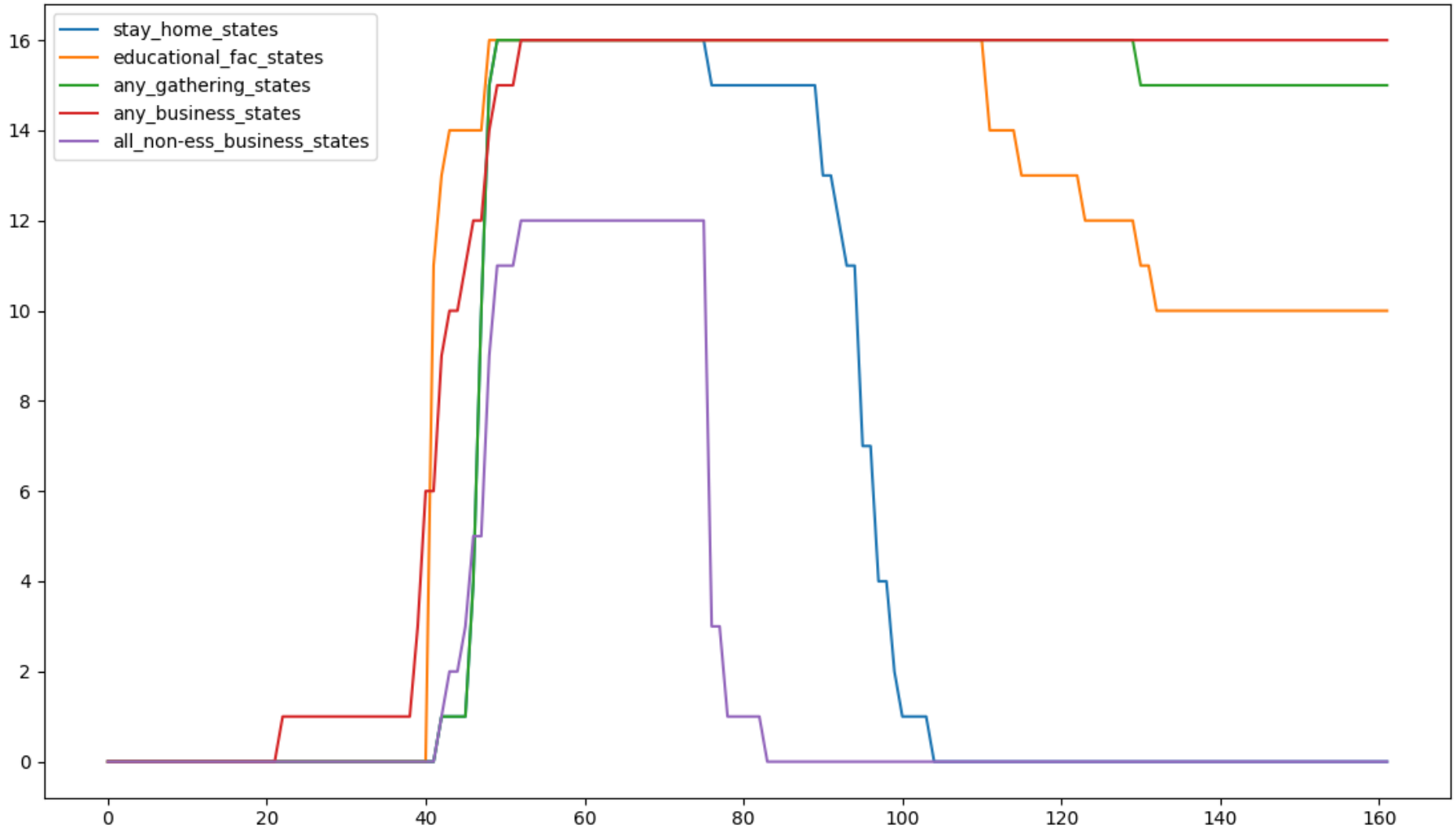
- Let's try to apply what we have learned to data sets about COVID-19, to understand the dynamics of the pandemic
  - Obvious: Only few experiments possible, urgent - perfect use case
- Here: Germany, Italy, US
- IHME data sets
  - Time series such as daily: Confirmed infections, hospitalization, ventilation capacity, mobility score, …
  - Combine with regional summary data provided by IHME

Regional Data for Germany - Augmentation

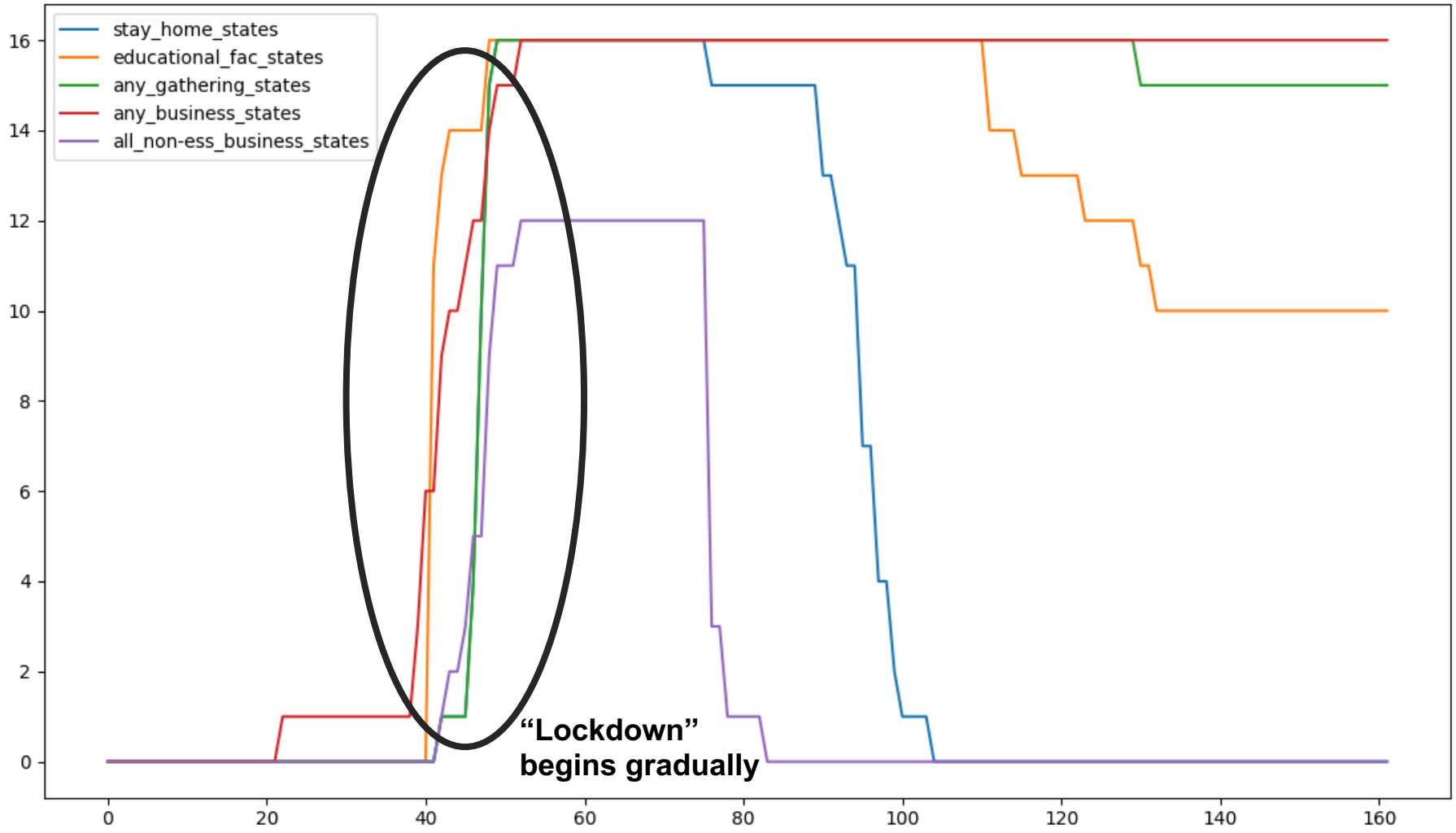
- Construct new time series that are indicators for interesting potential influences on the pandemic
- Federal state level ‘counters’: Increase when state applied measure, decrease when it is relaxed
  - Filter *manually* from all global IHME regions (result preview below)

location_name	stay_home_start_date	stay_home_end_date	educational_fac_start_date	educational_fac_end_date	any_gat
Mecklenburg-Vorpommern	2020-03-23	2020-05-07	2020-03-16		2020-03
Bremen	2020-03-22	2020-05-14	2020-03-16	2020-05-25	2020-03
Baden-Württemberg	2020-03-21	2020-05-11	2020-03-17		2020-03
North Rhine-Westphalia	2020-03-23	2020-05-11	2020-03-16		2020-03
Rhineland-Palatinate	2020-03-22	2020-05-13	2020-03-16		2020-03
Saxony	2020-03-23	2020-04-20	2020-03-23	2020-06-06	2020-03
Schleswig-Holstein	2020-03-24	2020-05-09	2020-03-16		2020-03
Saxony-Anhalt	2020-03-22	2020-05-04	2020-03-16	2020-06-15	2020-03
Lower Saxony	2020-03-23	2020-05-11	2020-03-16		2020-03
Brandenburg	2020-03-17	2020-05-09	2020-03-18	2020-05-25	2020-03
Berlin	2020-03-23	2020-05-09	2020-03-23	2020-05-29	2020-03
Bavaria	2020-03-21	2020-05-06	2020-03-16		2020-03
Saarland	2020-03-21	2020-05-18	2020-03-16		2020-03

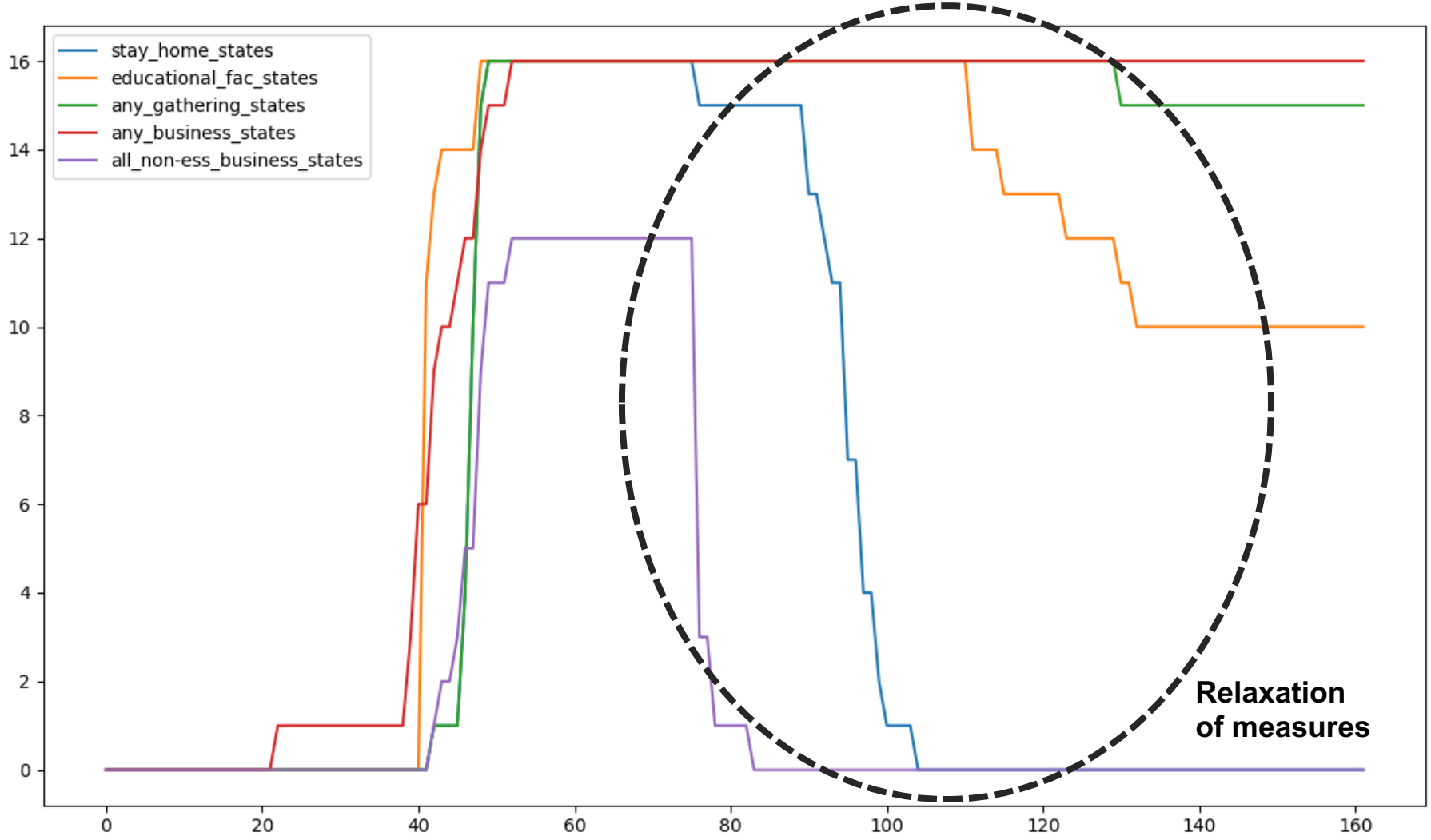
Regional Data for Germany: Time Series Added



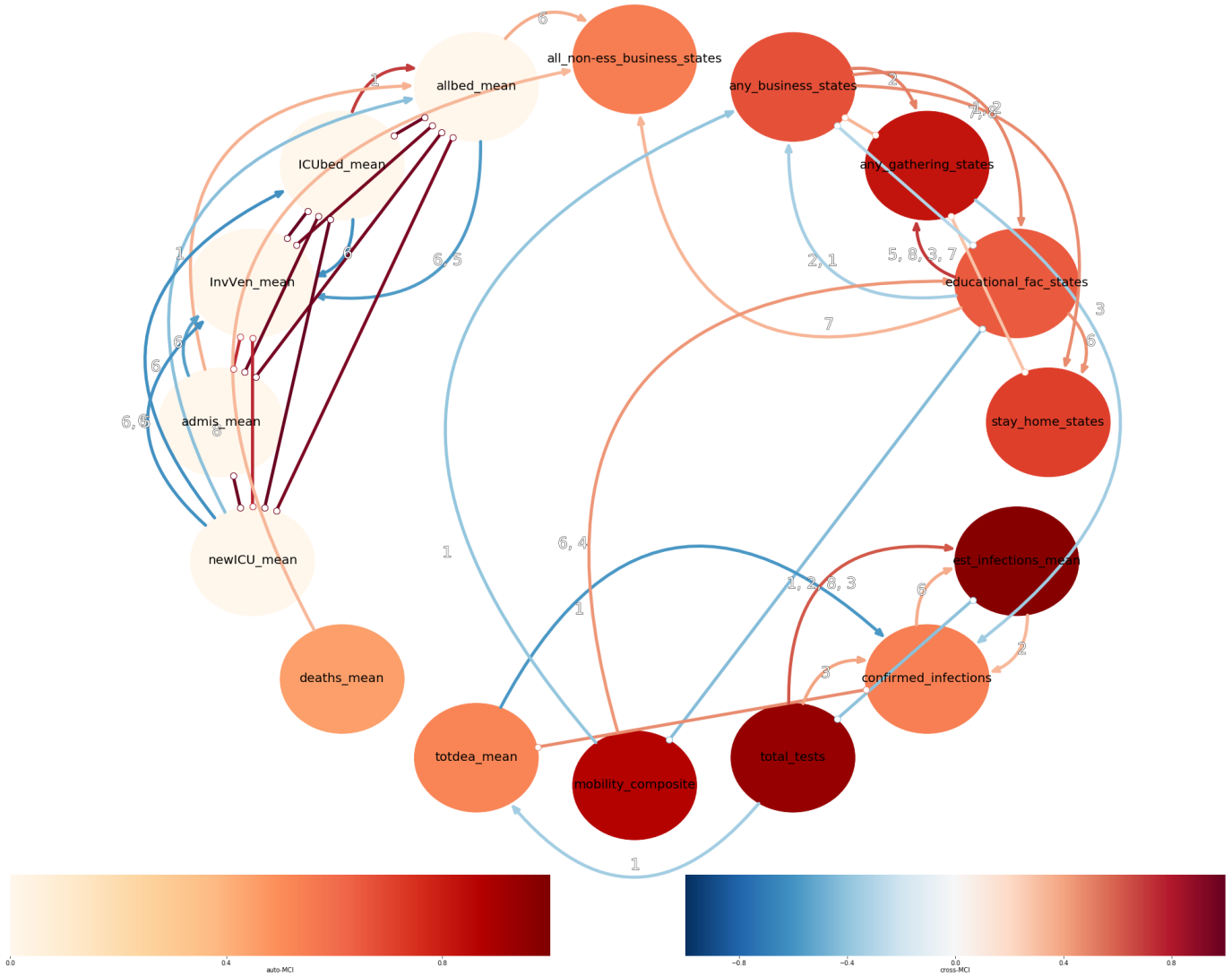
Regional Data for Germany: Time Series Added



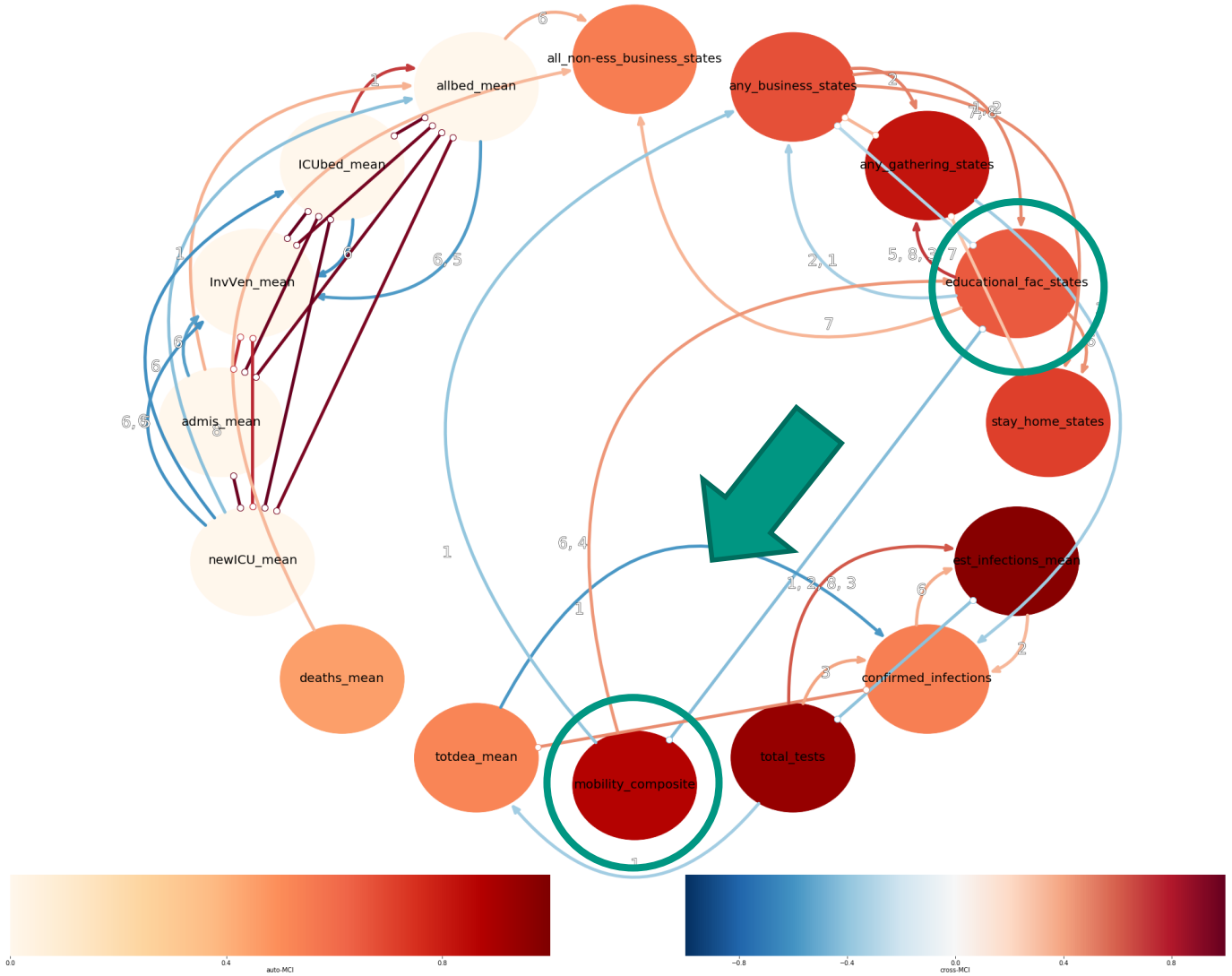
Regional Data for Germany: Time Series Added



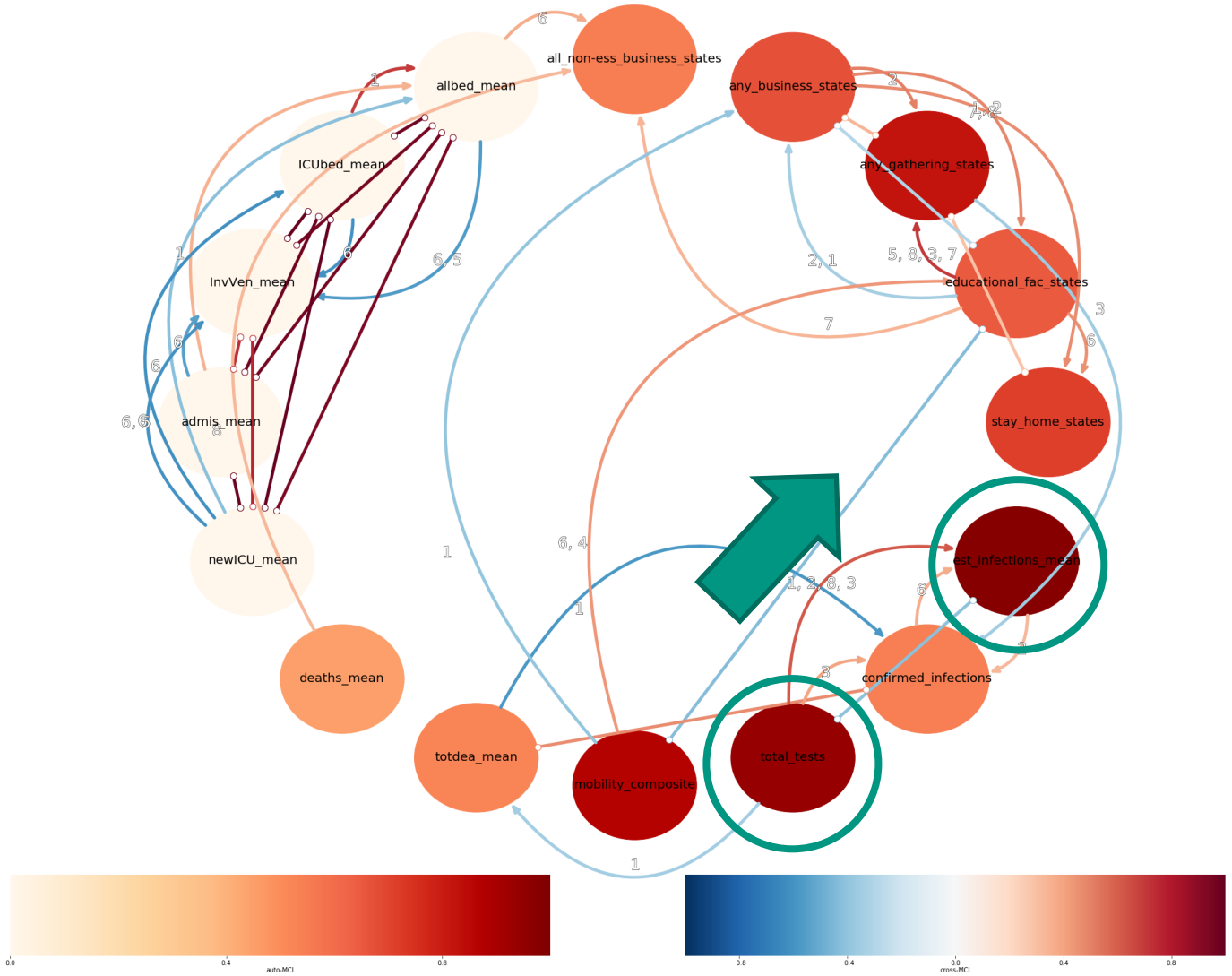
PCMCI for COVID-19: Germany (Tau=8)



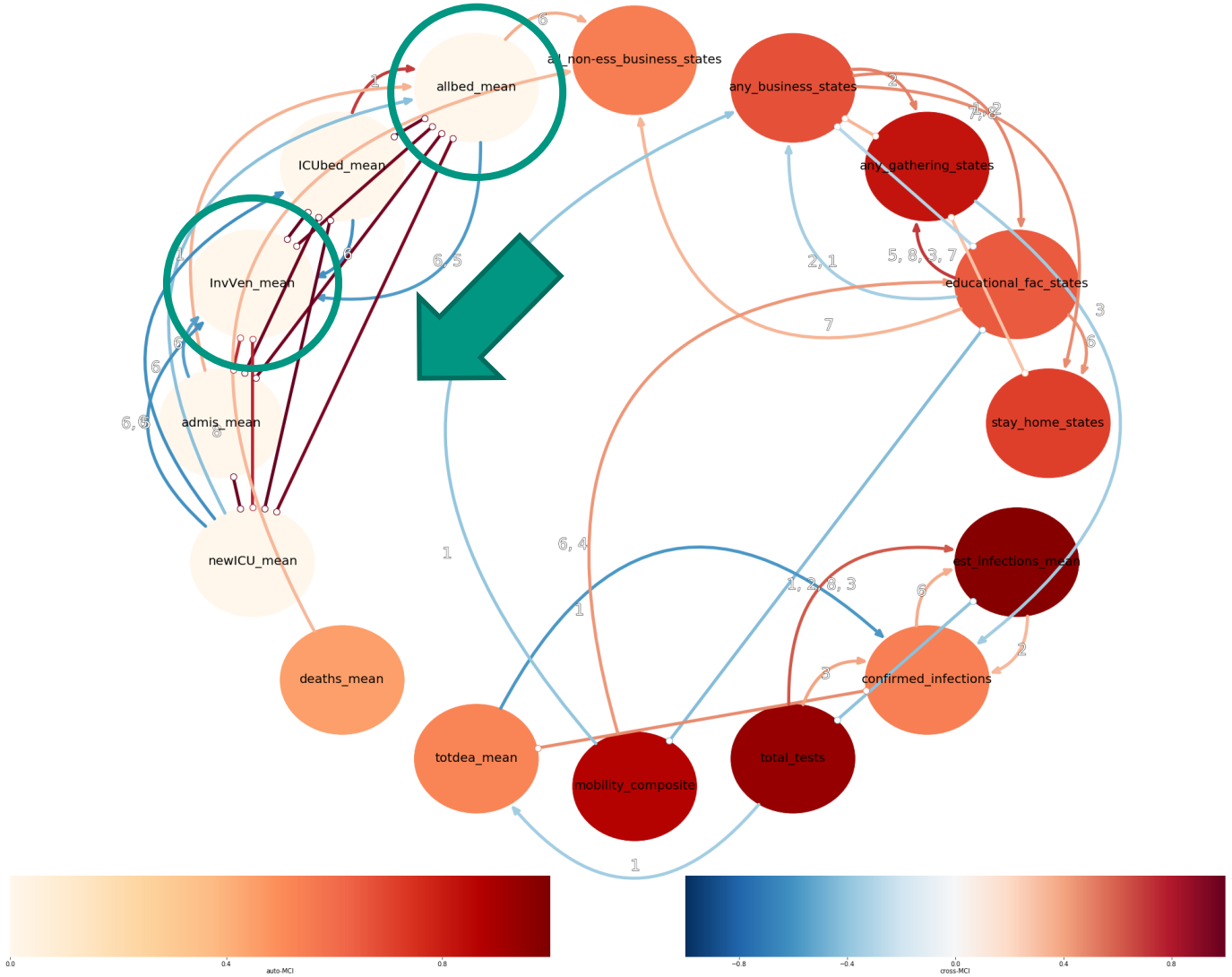
PCMCI for COVID-19: Germany (Tau=8)



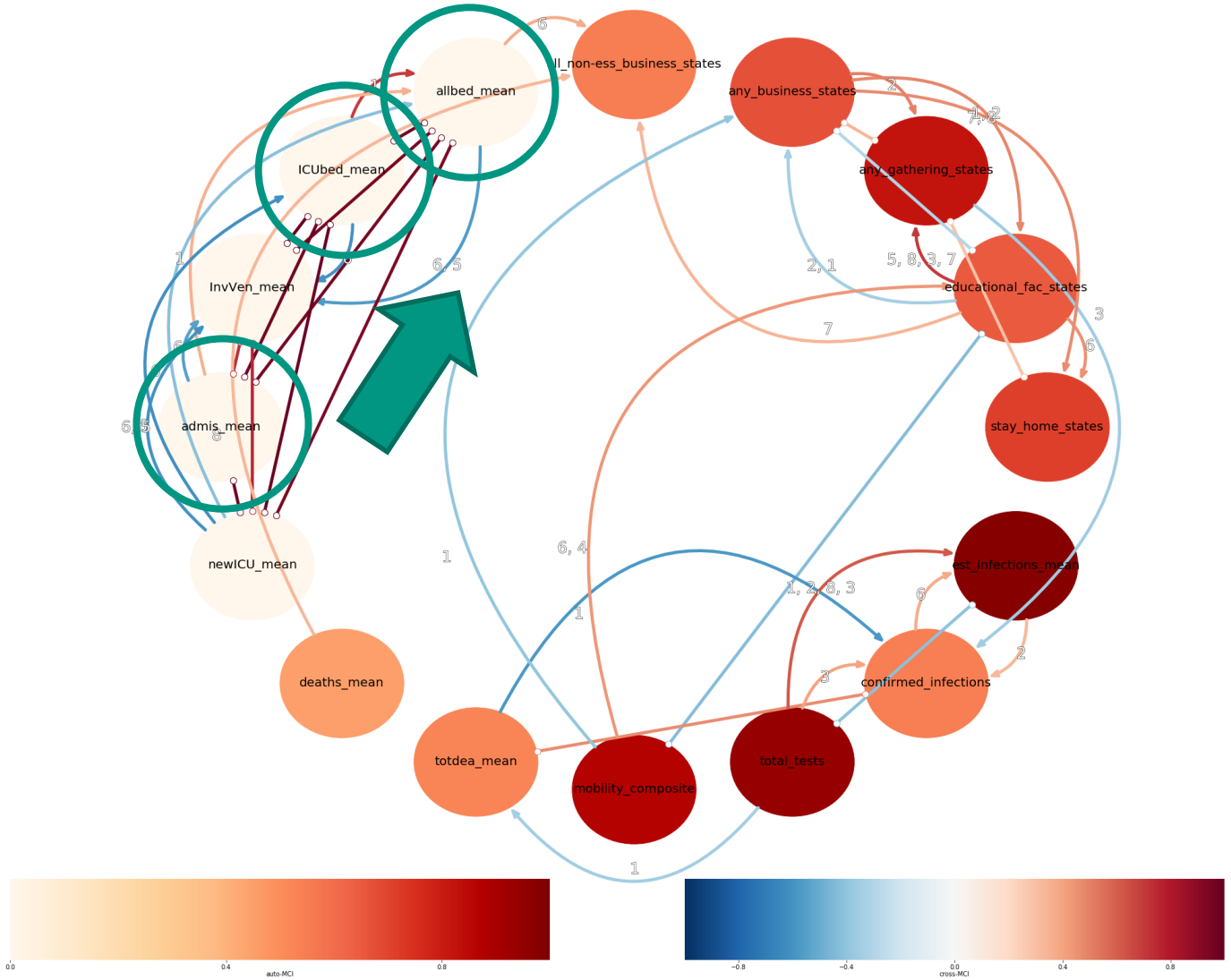
PCMCI for COVID-19: Germany (Tau=8)



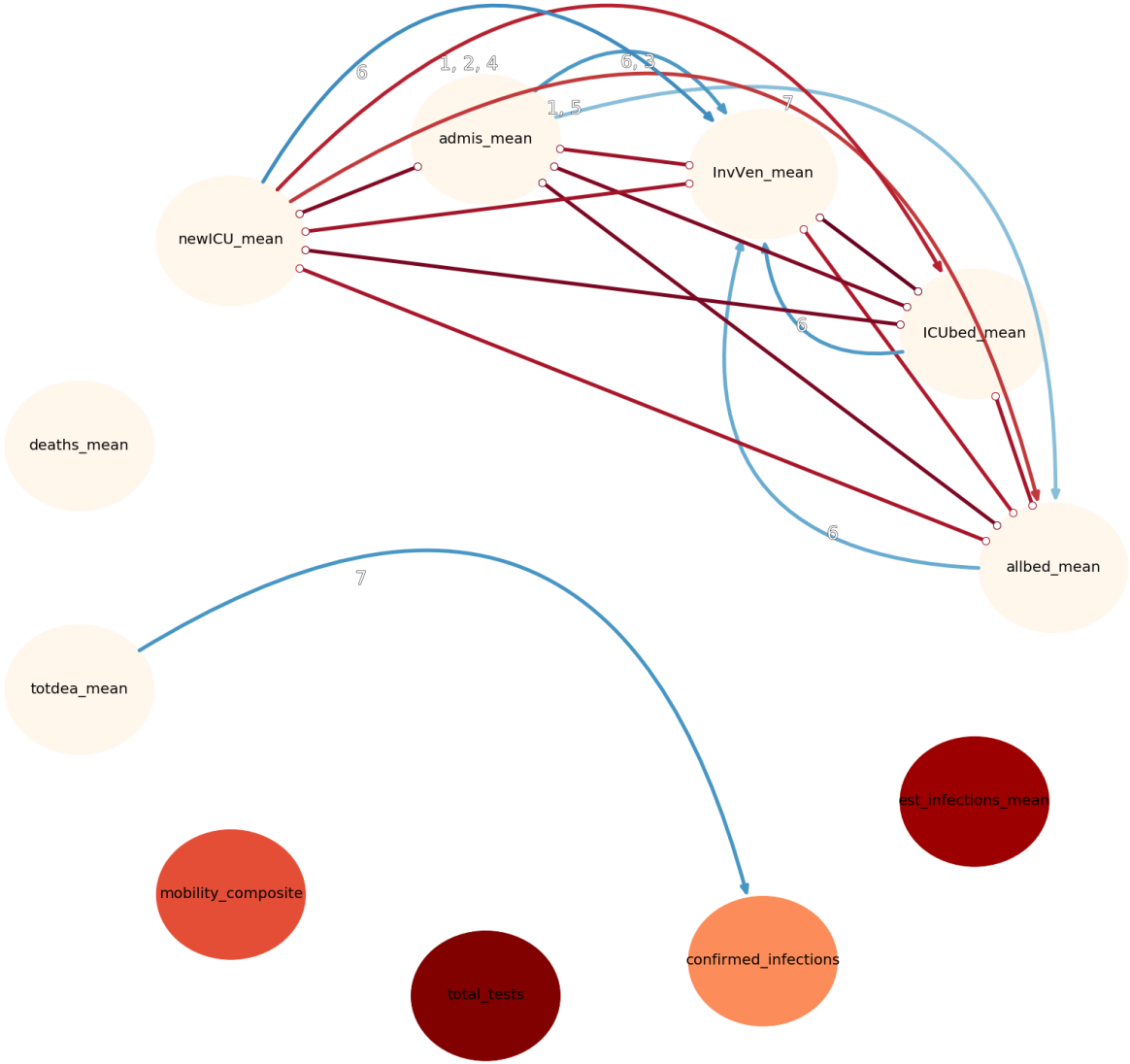
PCMCI for COVID-19: Germany (Tau=8)



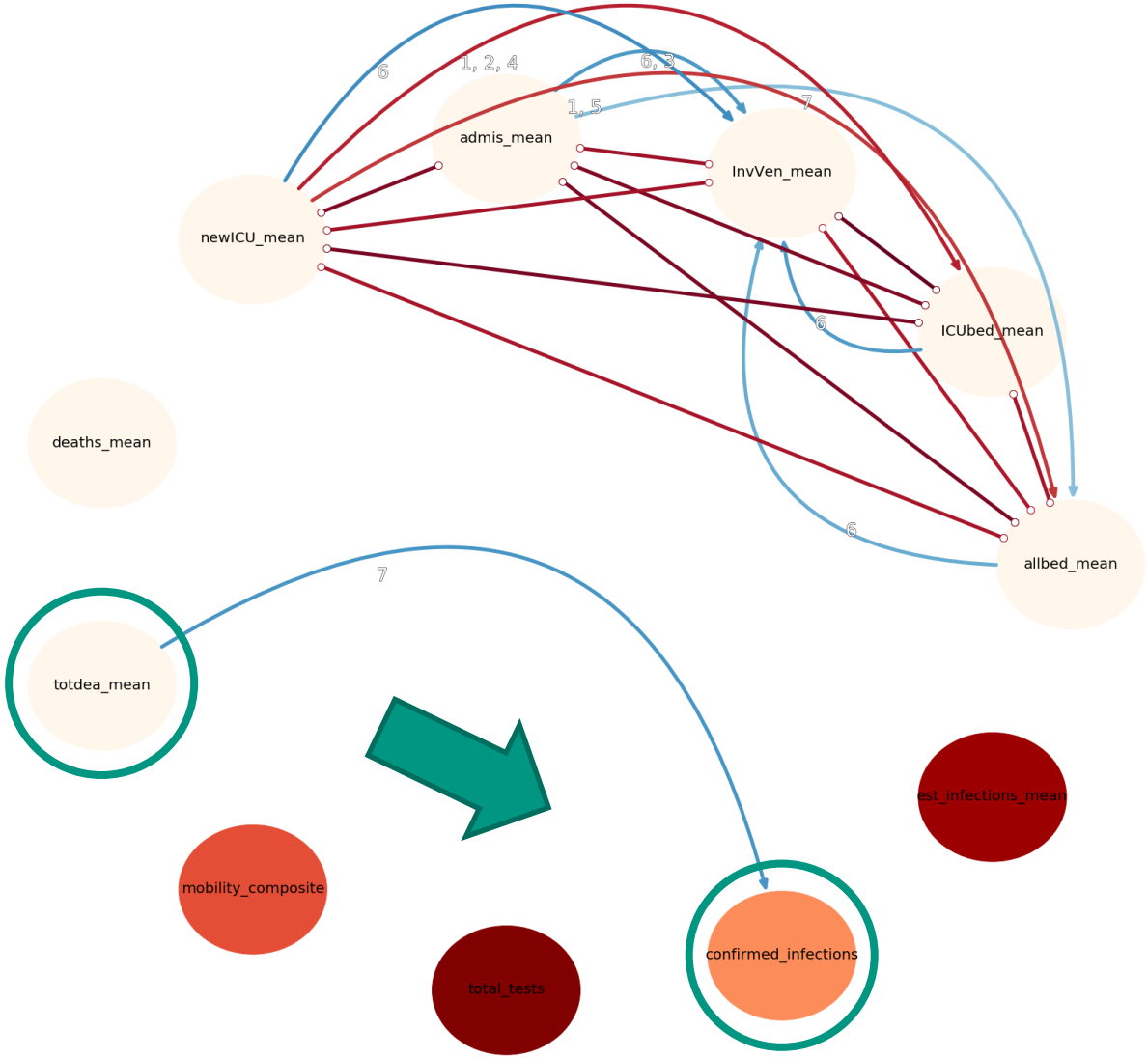
PCMCI for COVID-19: Germany (Tau=8)



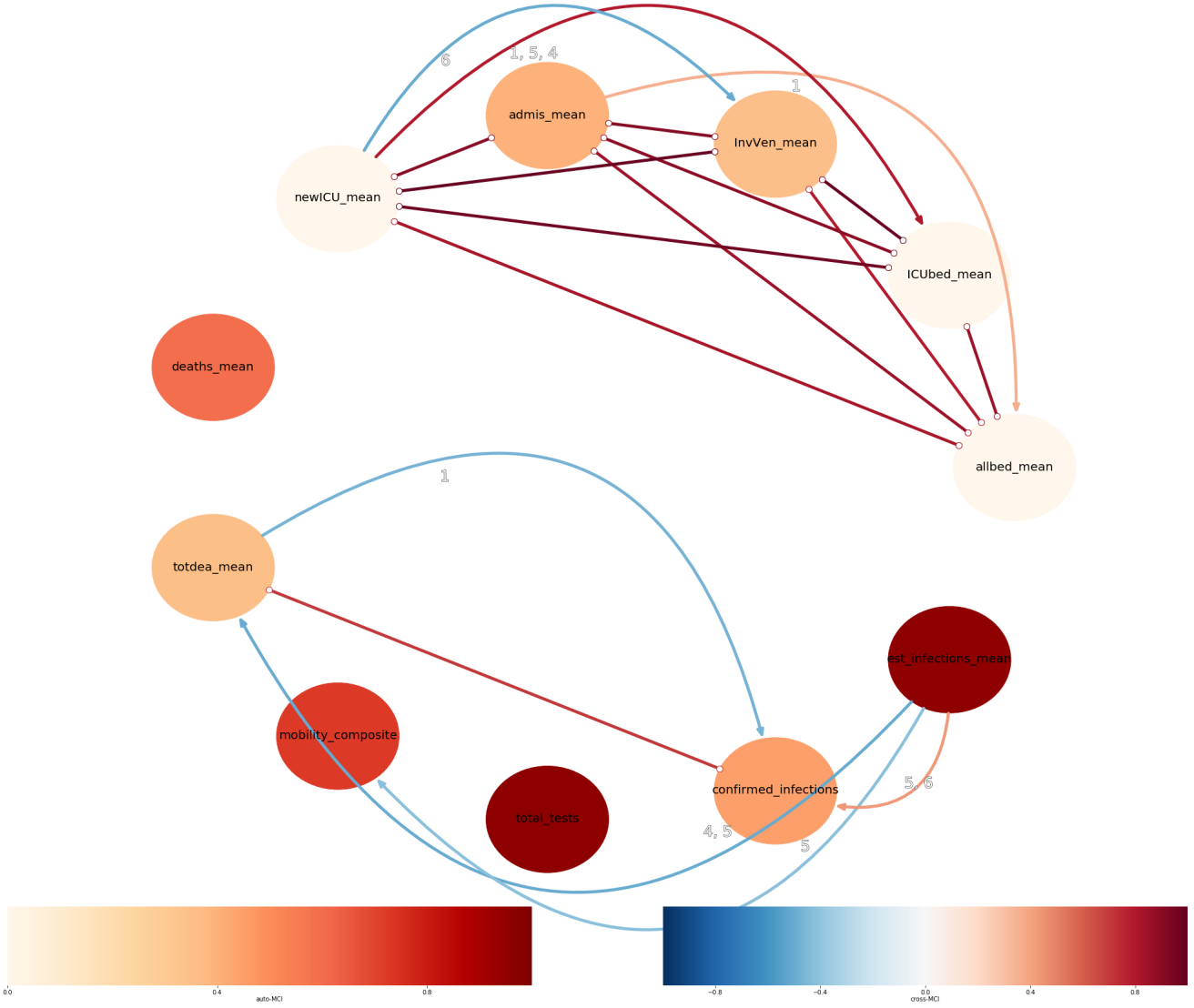
PCMCI for COVID-19: Italy (Tau=8)



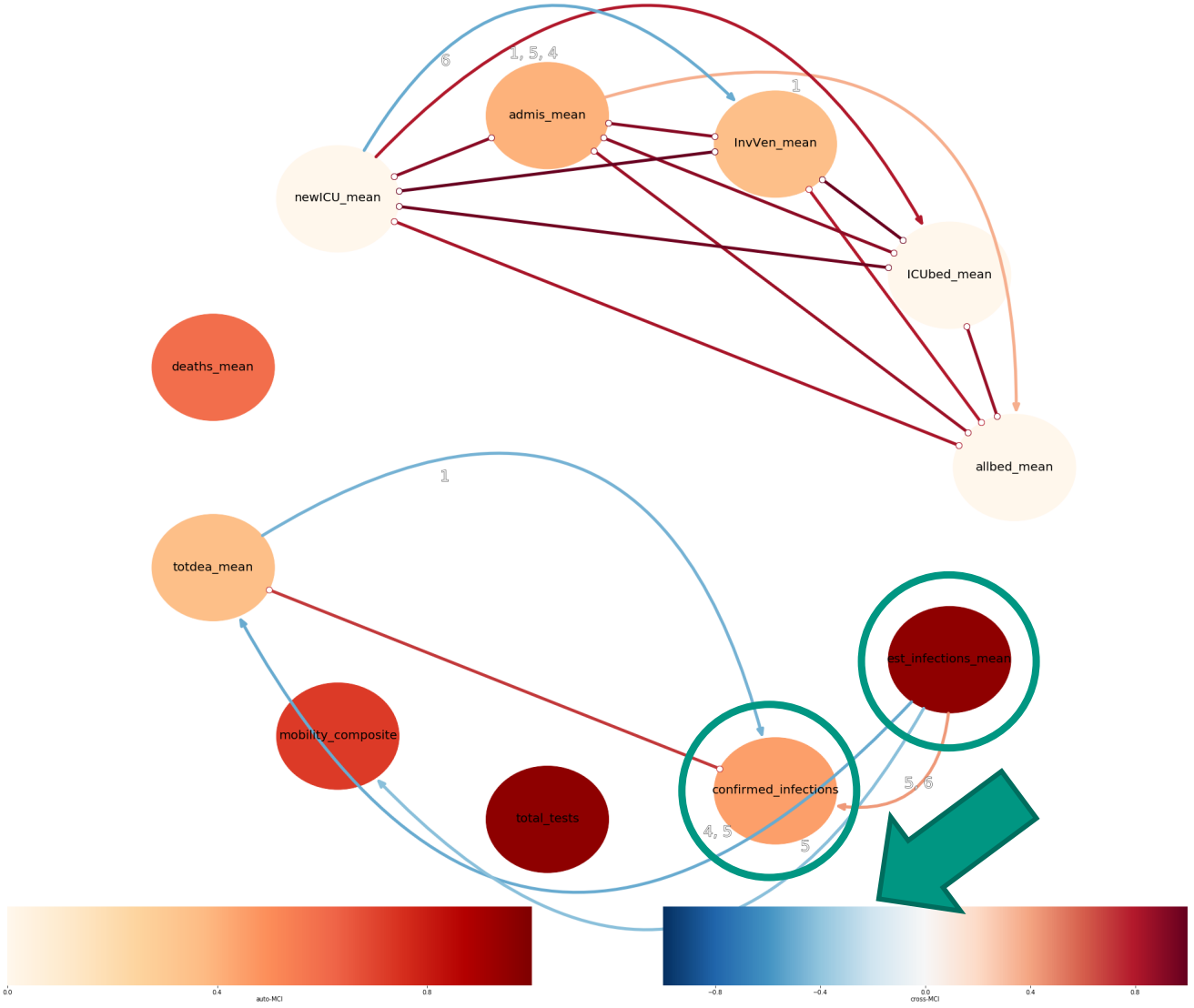
PCMCI for COVID-19: Italy (Tau=8)



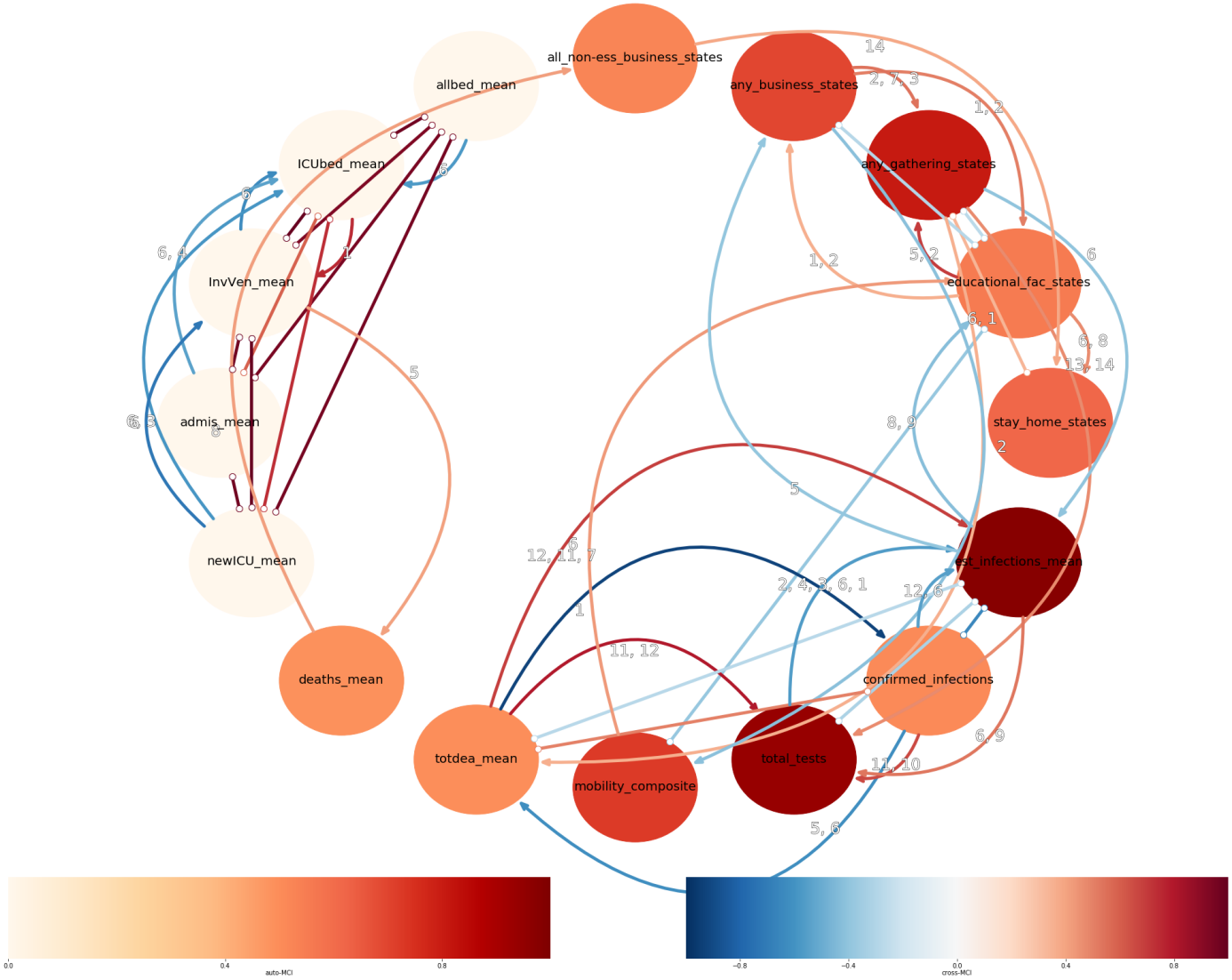
PCMCI for COVID-19: US (Tau=8)



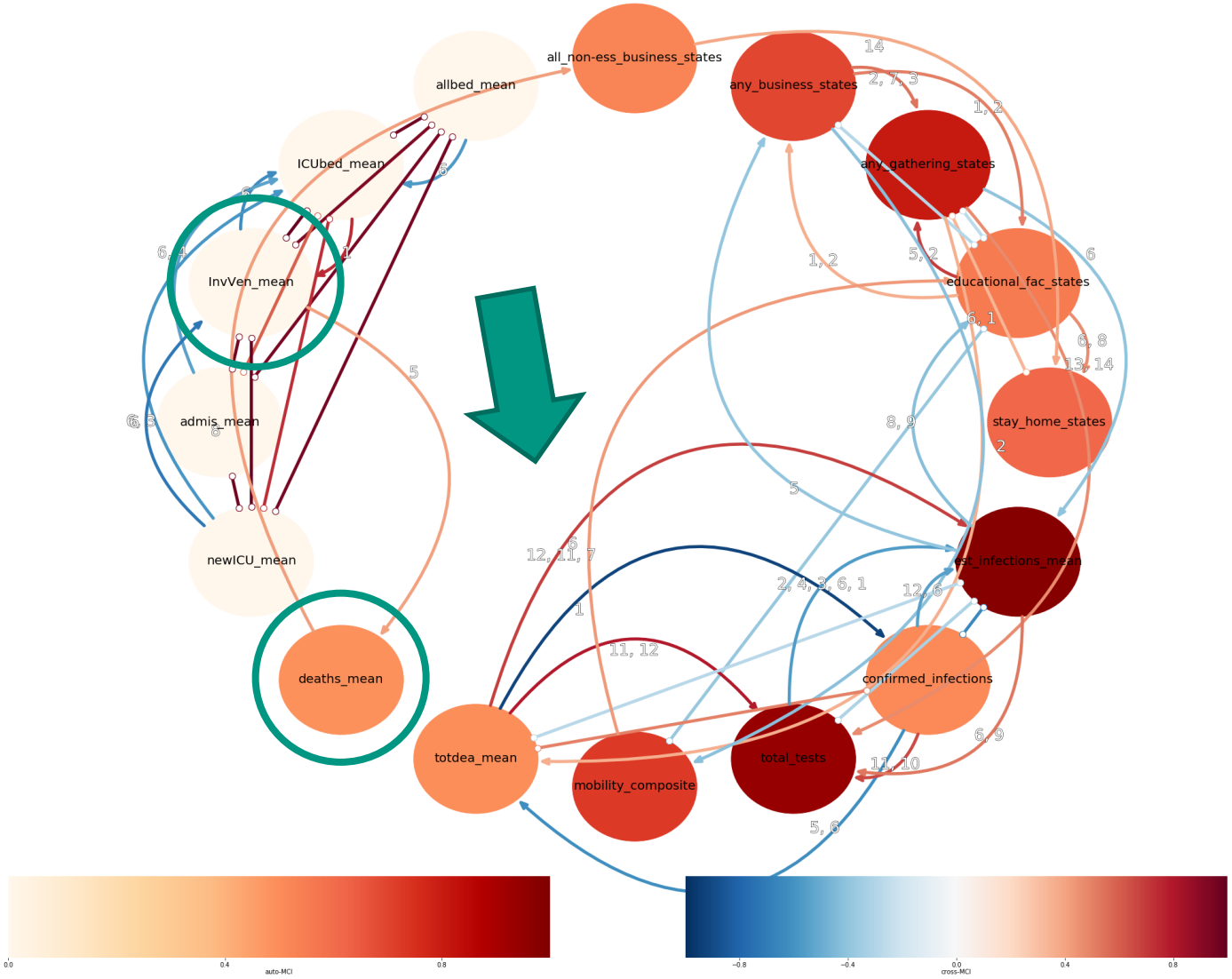
PCMCI for COVID-19: US (Tau=8)



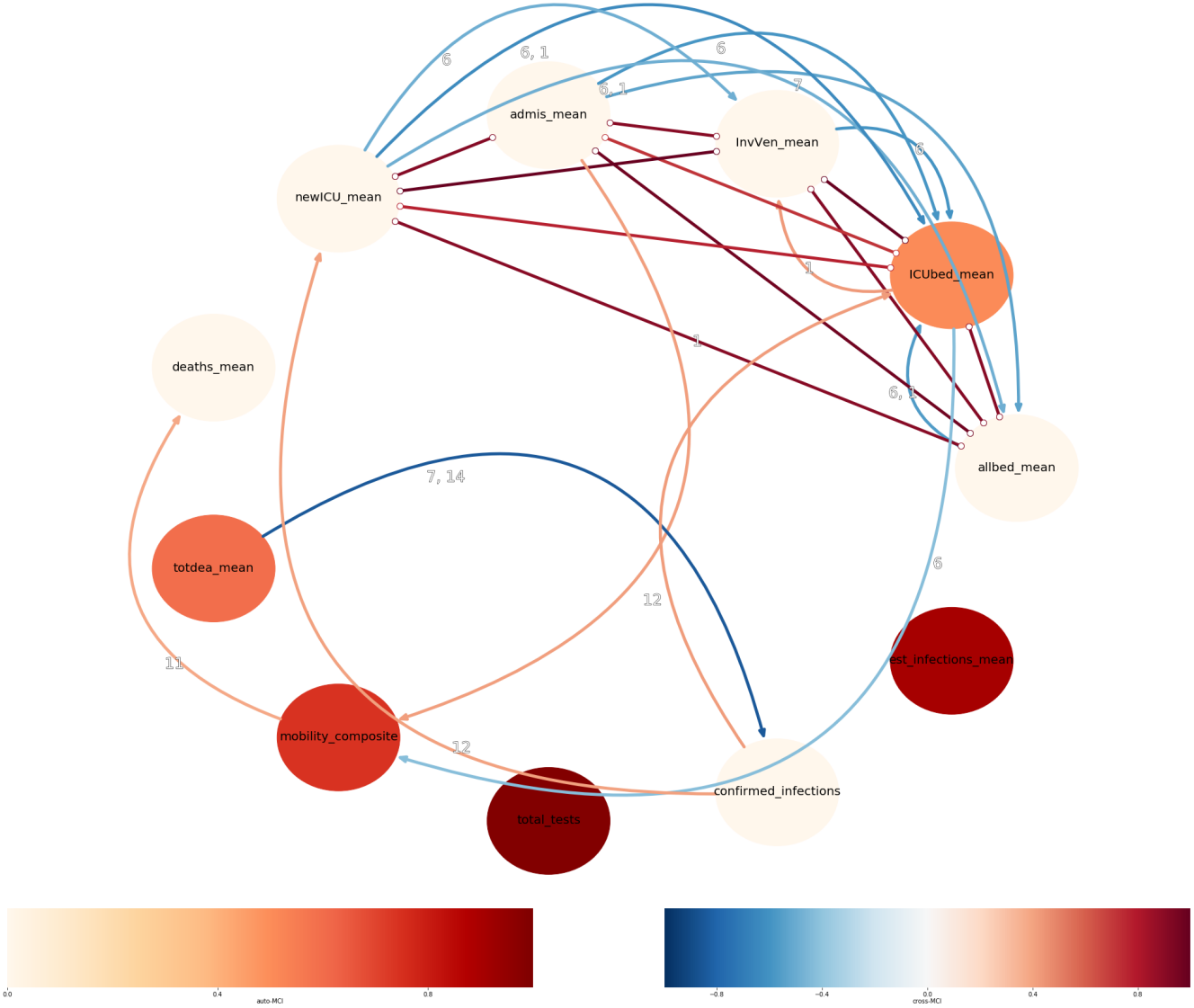
PCMCI for COVID-19: Germany (Tau=14)



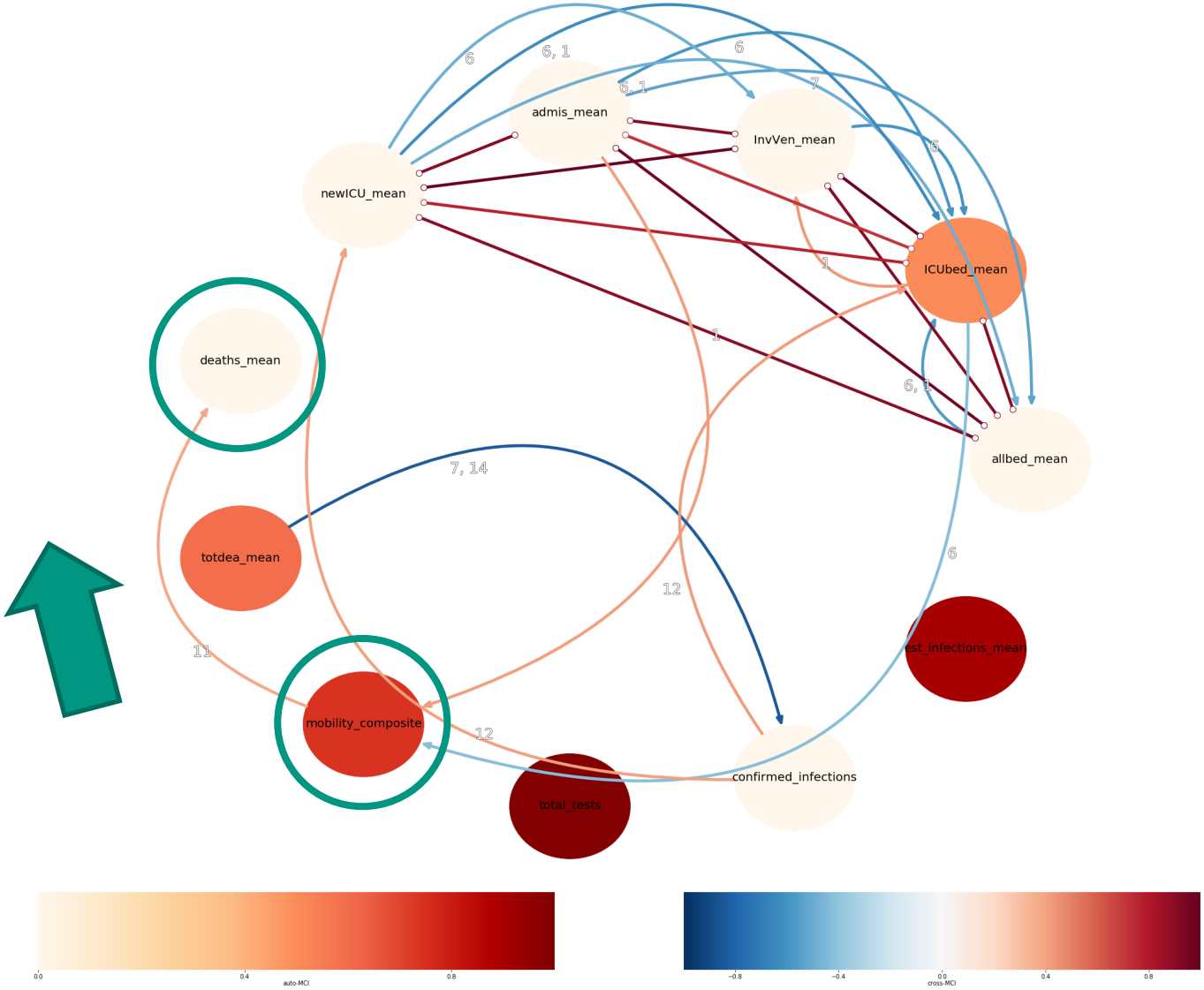
PCMCI for COVID-19: Germany (Tau=14)



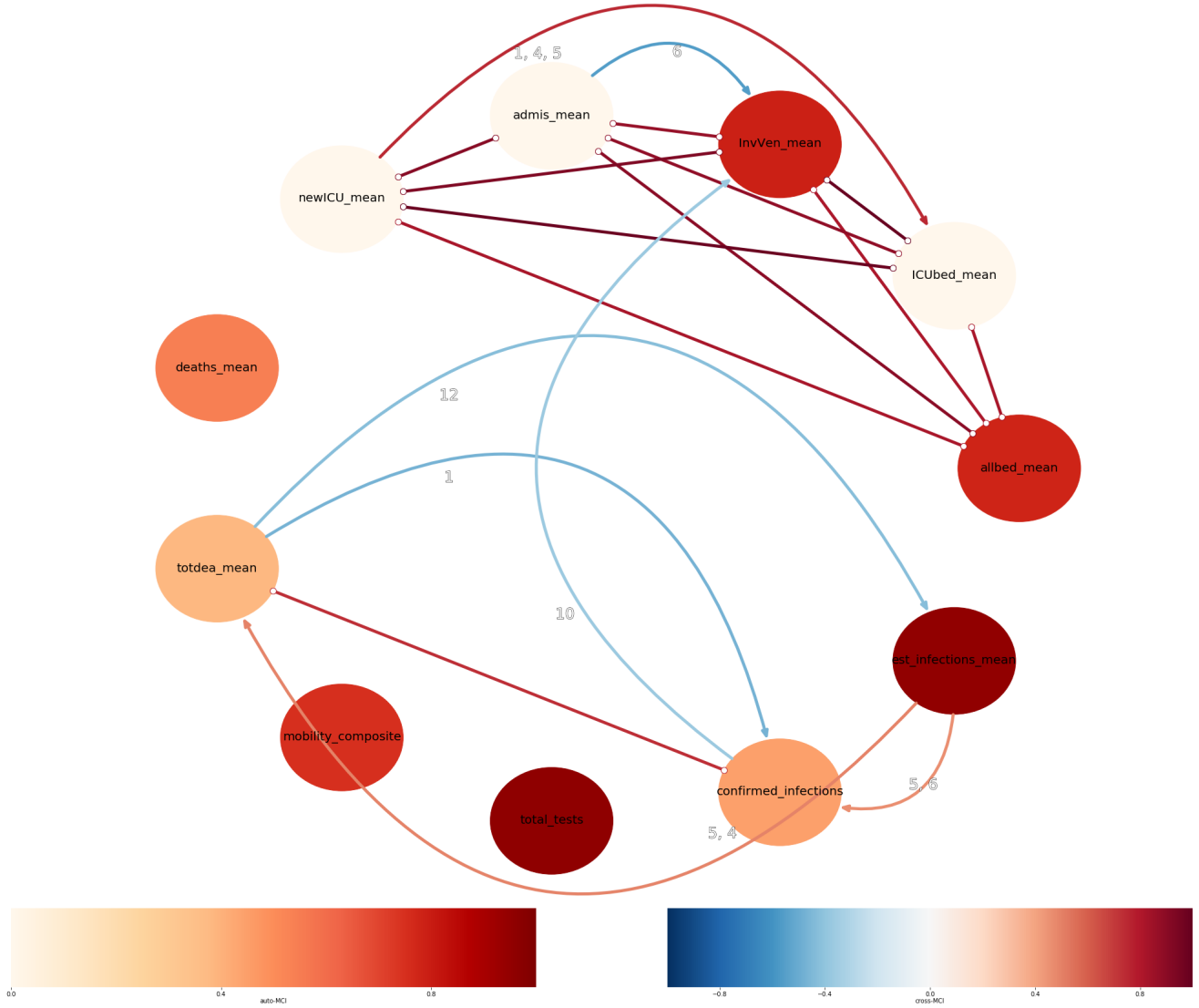
PCMCI for COVID-19: Italy (Tau=14)



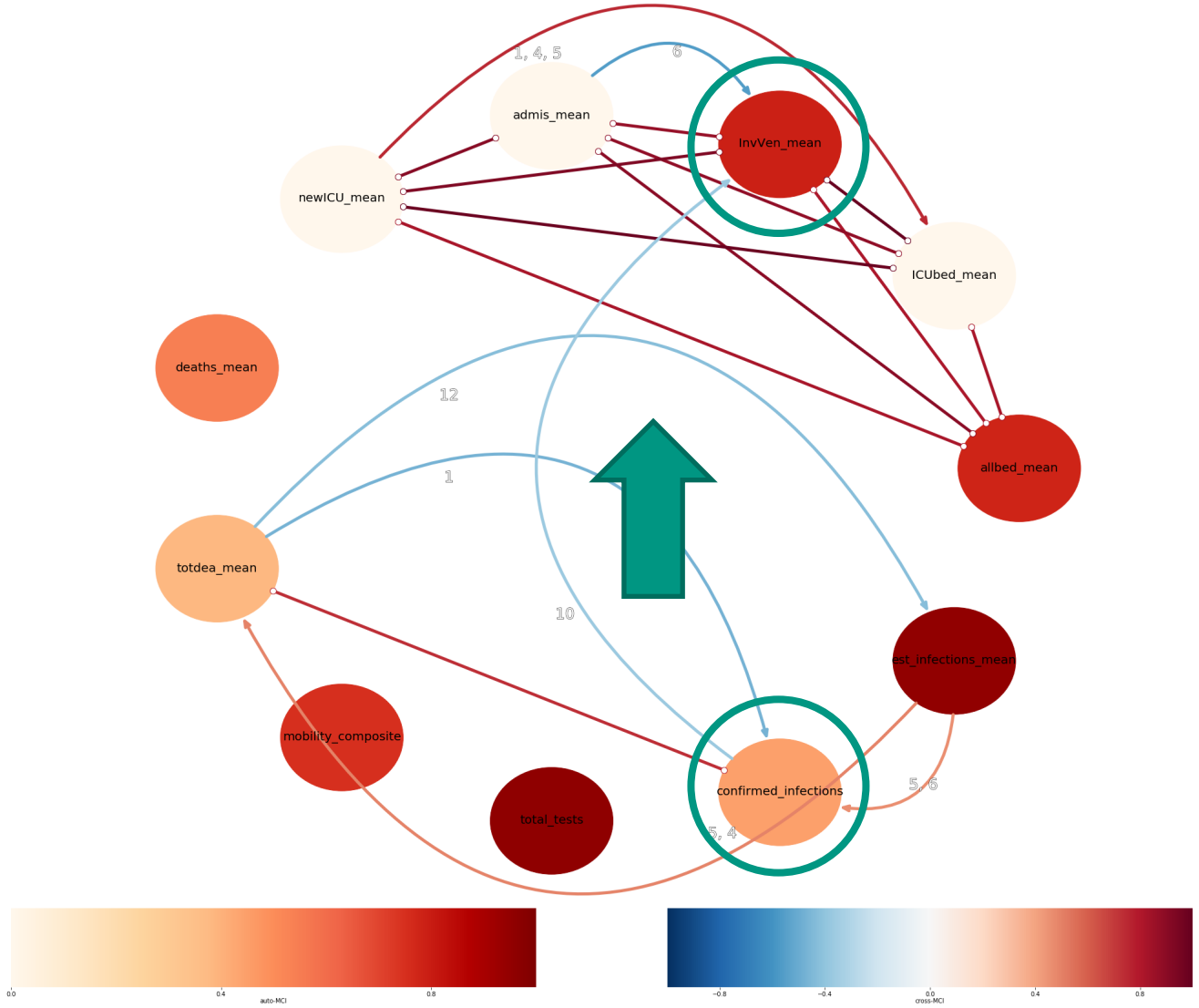
PCMCI for COVID-19: Italy (Tau=14)



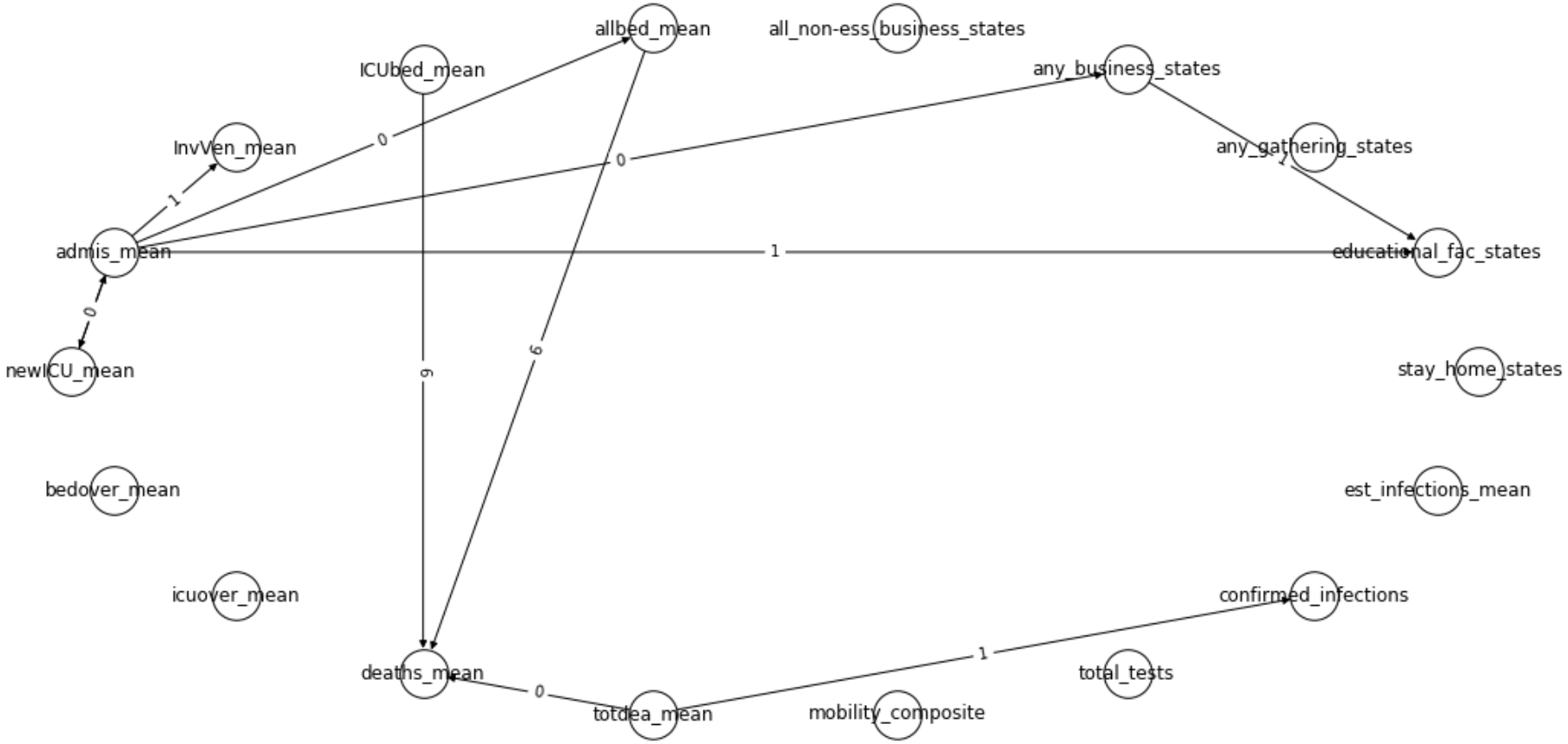
PCMCI for COVID-19: US (Tau=14)



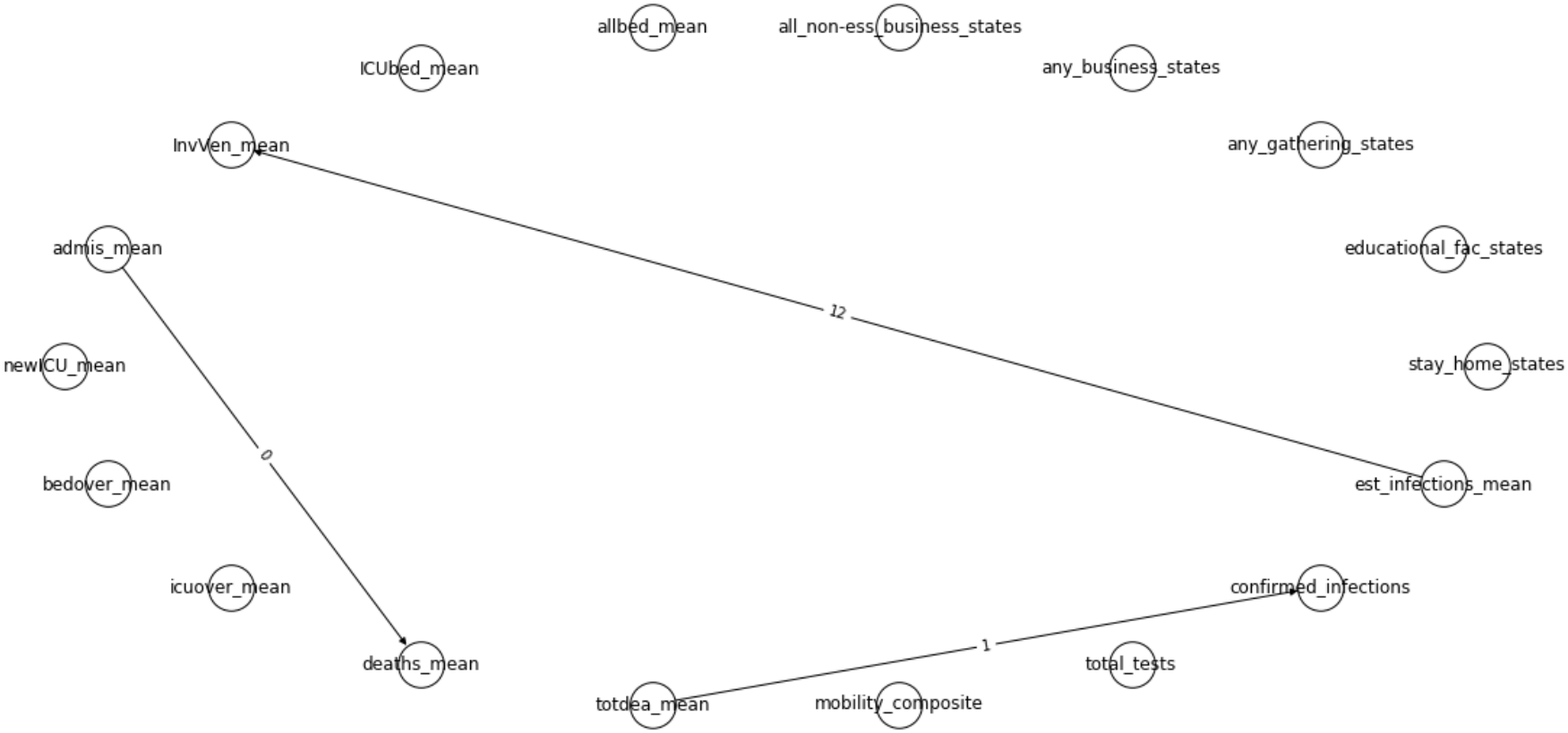
PCMCI for COVID-19: US (Tau=14)



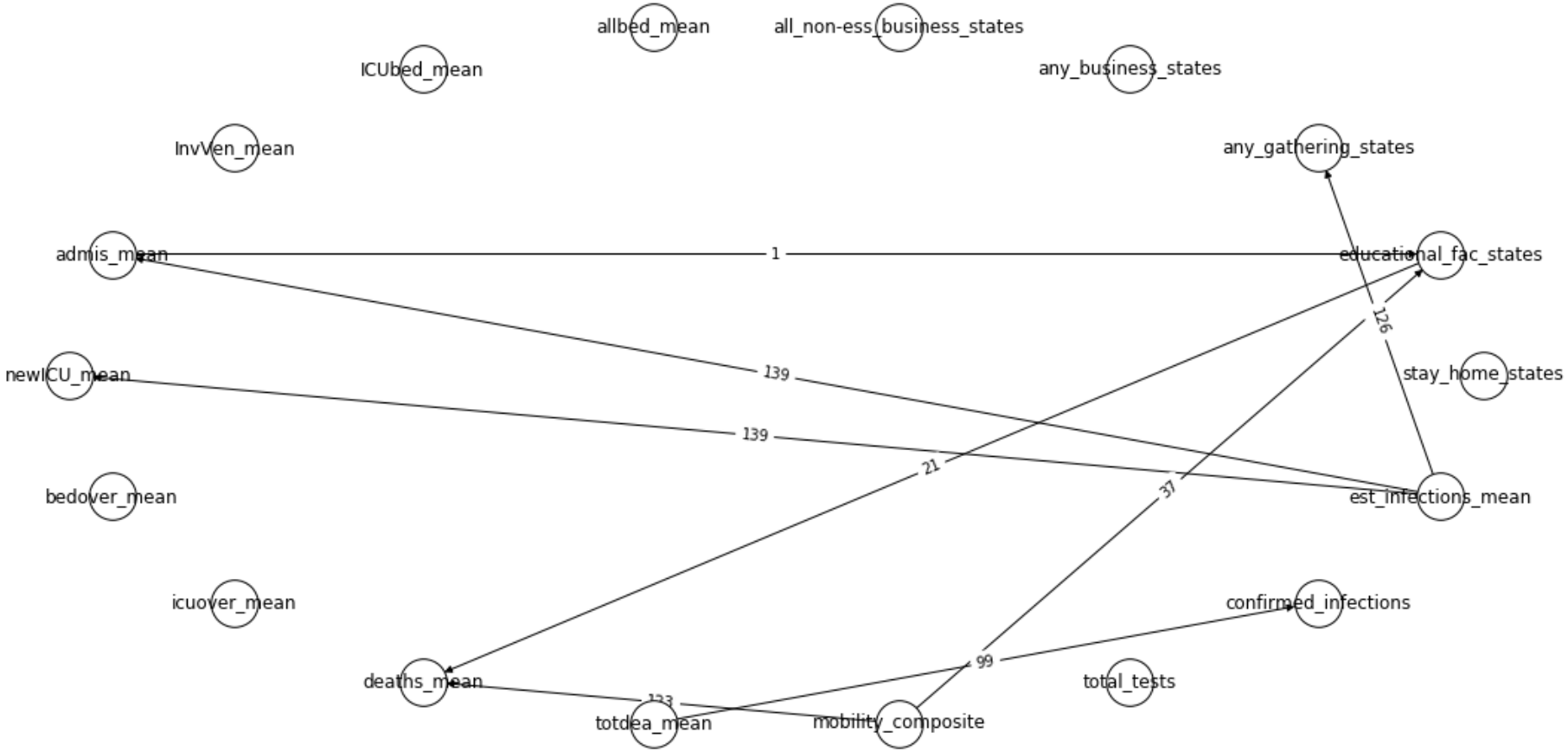
TCDF for COVID-19: Germany



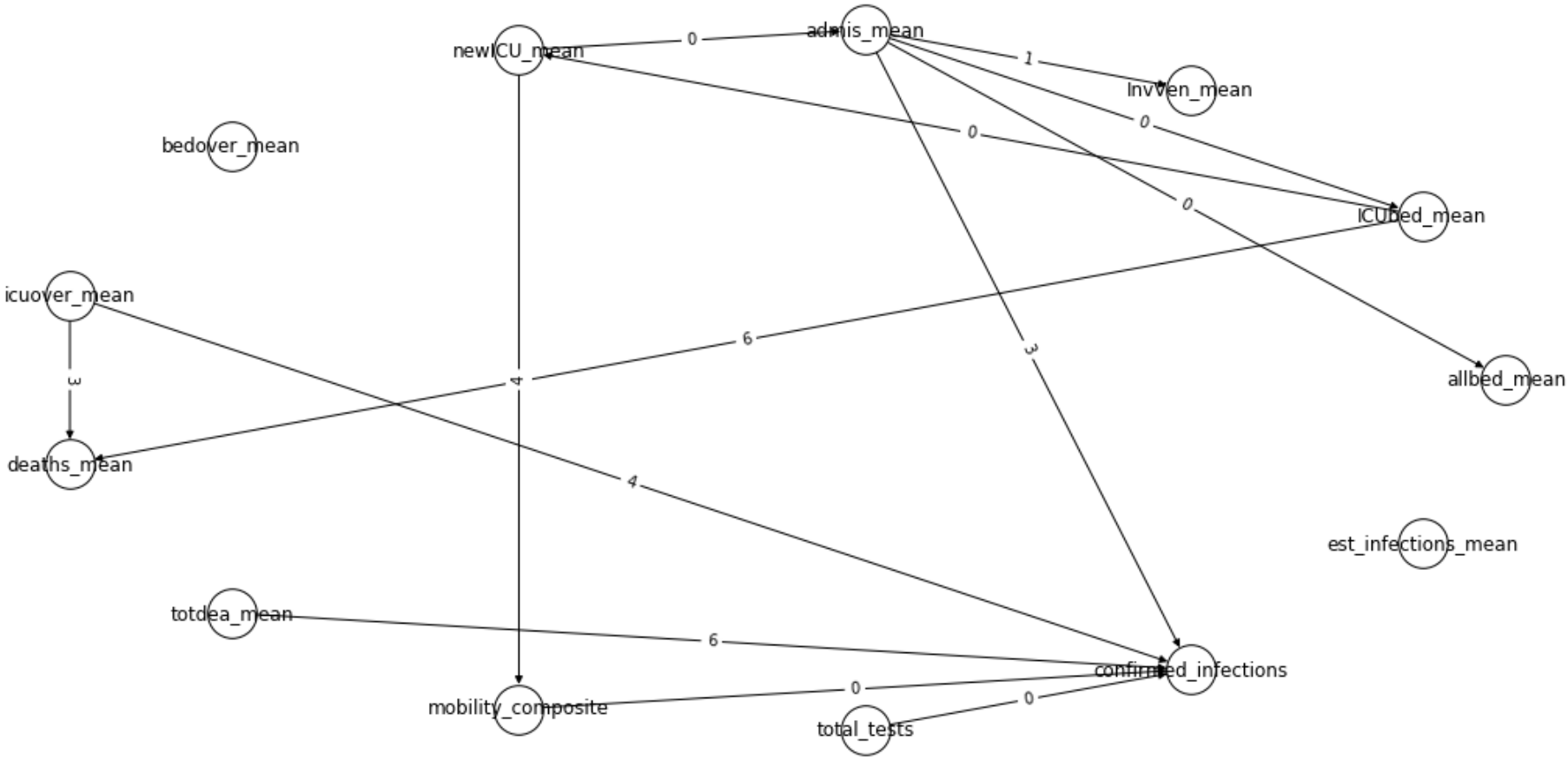
TCDF for COVID-19: Germany (kernels=14)



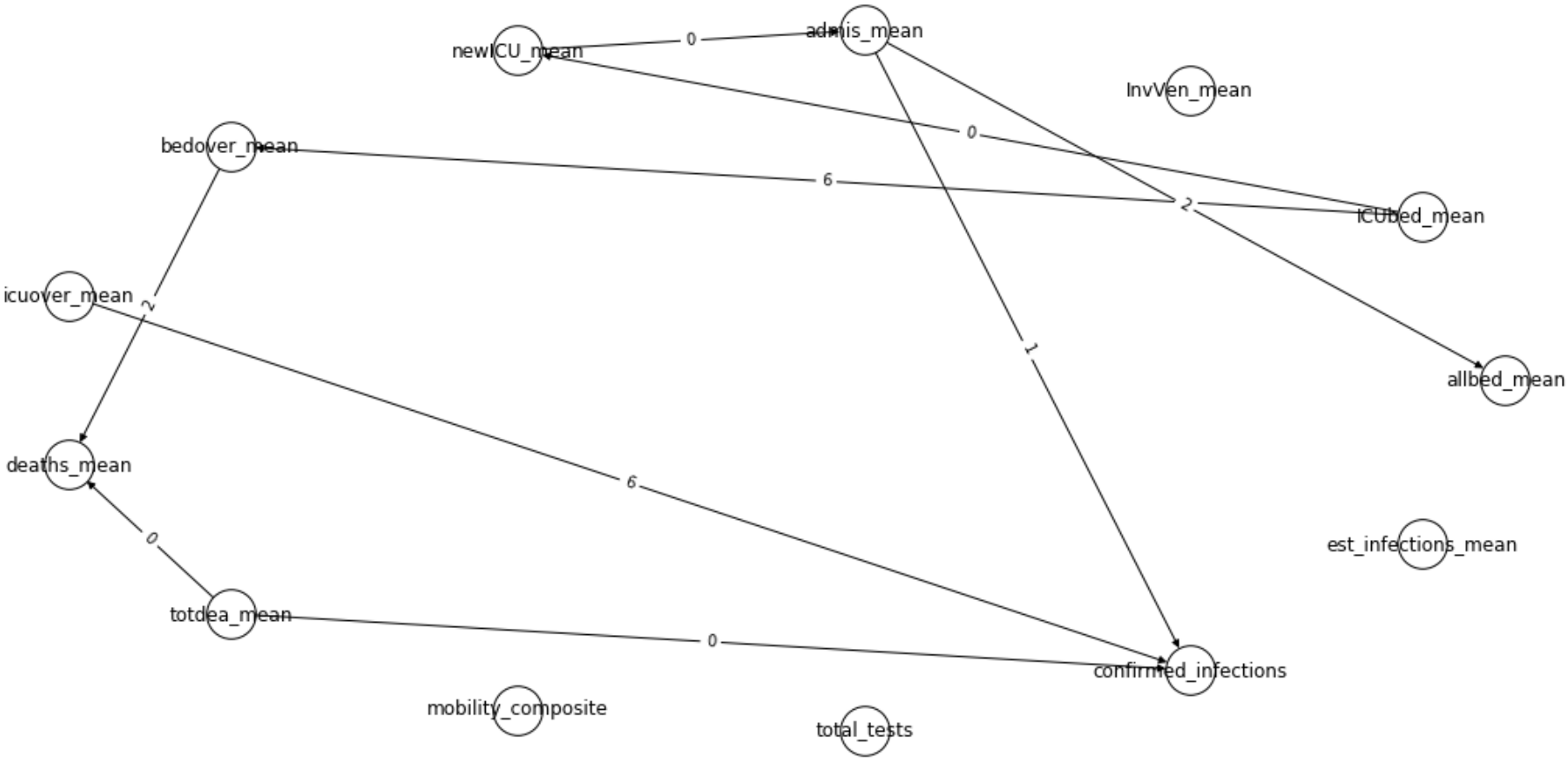
TCDF for COVID-19: Germany (kernels=14, one hidden layer)



TCDF for COVID-19: Italy (lr=0.01)



TCDF for COVID-19: US (lr=0.01)

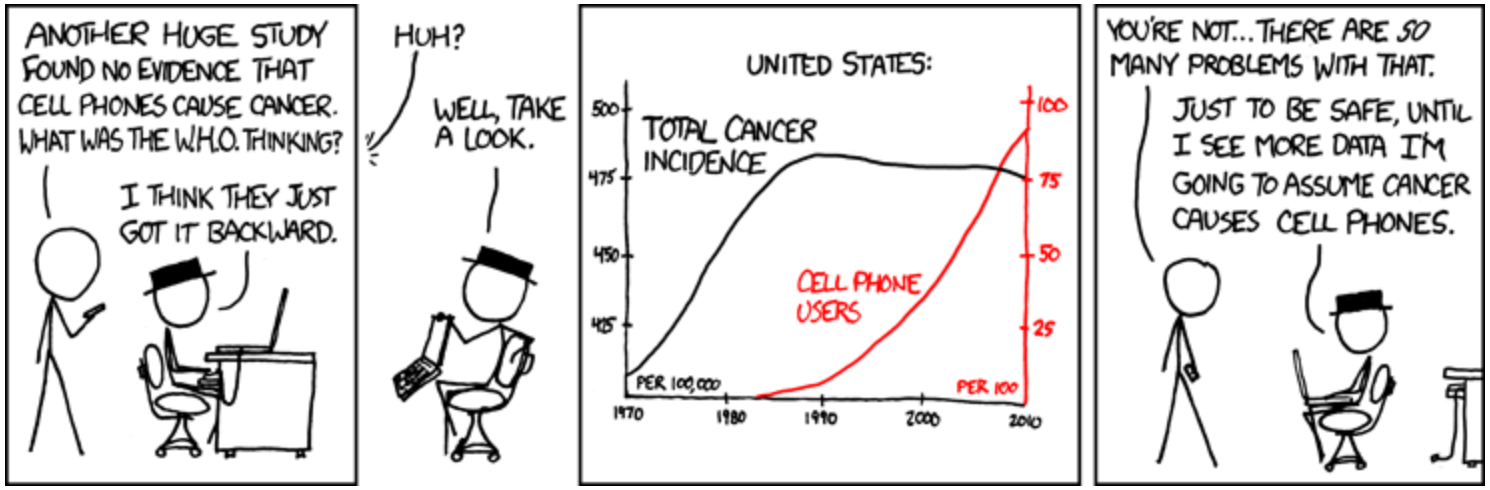


Summary & Conclusions

- Many interesting application scenarios supporting scientific research for causal inference and discovery
- Today, many limitations
- Validation of unsupervised methods can be challenging – careful approach and domain knowledge often needed

- Causal Inference a future “hot topic” for ML research

Thank you for your attention!



https://xkcd.com/925/

PCMCI Exploration (Backup)

- Very good performance on basic linear datasets
- Runtime can increase heavily when increasing variables or time steps
- Worse performance on other datasets
- Runtime very high for certain experiments