A homework solution guide for decision support with causal Bayesian networks

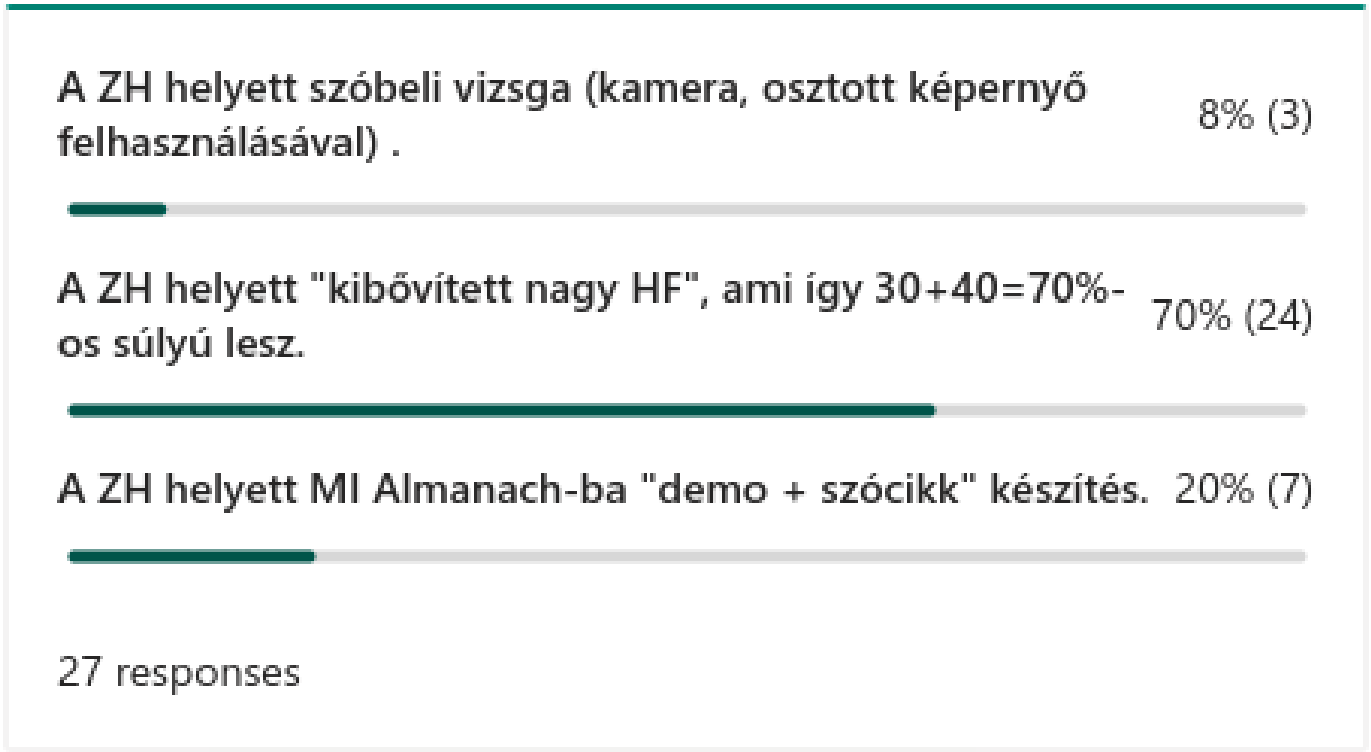
Péter Antal

Outline

- Midterm in distance education: poll
  - Oral exam, extended homework, "AI Almanach"
- Goals of the homework
- Earlier solutions
- Candidate domains and workgroups(!)
- Tasks in the homework
  - The default part and the midterm extension
- "AI Almanach" as second homework

Midterm in distance education

- Poll in Teams PDSS.general channel:



Goals of the homework

To demonstrate and practice multifaceted nature of Bayesian networks (BNs)

- As a probabilistic logic knowledge base, it provides a coherent framework to represents beliefs.
- As a decision network, it provides a coherent framework to represent preferences for actions.
- As a dependency map, it explicitly represents the system of conditional independencies in a given domain.
- As a causal map, it explicitly represents the system of causal relations in a given domain.

==> As a decomposable probabilistic graphical model, it parsimoniously represents the quantitative stochastic dependencies of a domain and it allows efficient **observational inference**.

==> As an uncertain causal model, it parsimoniously represents the quantitative, stochastic, autonomous mechanisms in a domain and it allows efficient **interventional and counterfactual inference**.

The **default** and midterm parts

- **Default part (7 subtasks):**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Quantify your BN model.
  - Check it with global inference and „information sensitivity of inference" analysis.
  - Check it by relearning it from self-generated data.
  - Demonstrate observational, causal, and counterfactual inference in the model.
  - Extend your BN model to a decision network.
  - Investigate the value of further information.
- Midterm part (4 subtasks):

– Write a formal specification for your model with test cases.

- Perform ALL(!) the subtasks in the default homework using pomegranate
- Perform and document additional steps either in BayesCube or pomegranate
  - Analyse estimation biases
  - Investigate the effect of model uncertainty and sample size on learning.

Earlier homework topics

- Travel: how to travel to the university
- Education: personal performance
- Mental health: burn-out, learning attitude
- Tech: fault discovery (PC, mobile)
- Customer: how to buy (laptop, mobile)
- Software engineering: platform selection
- Existential: alone? threats?
- Misc.: dogs, cats, aquarium, sailing boat....

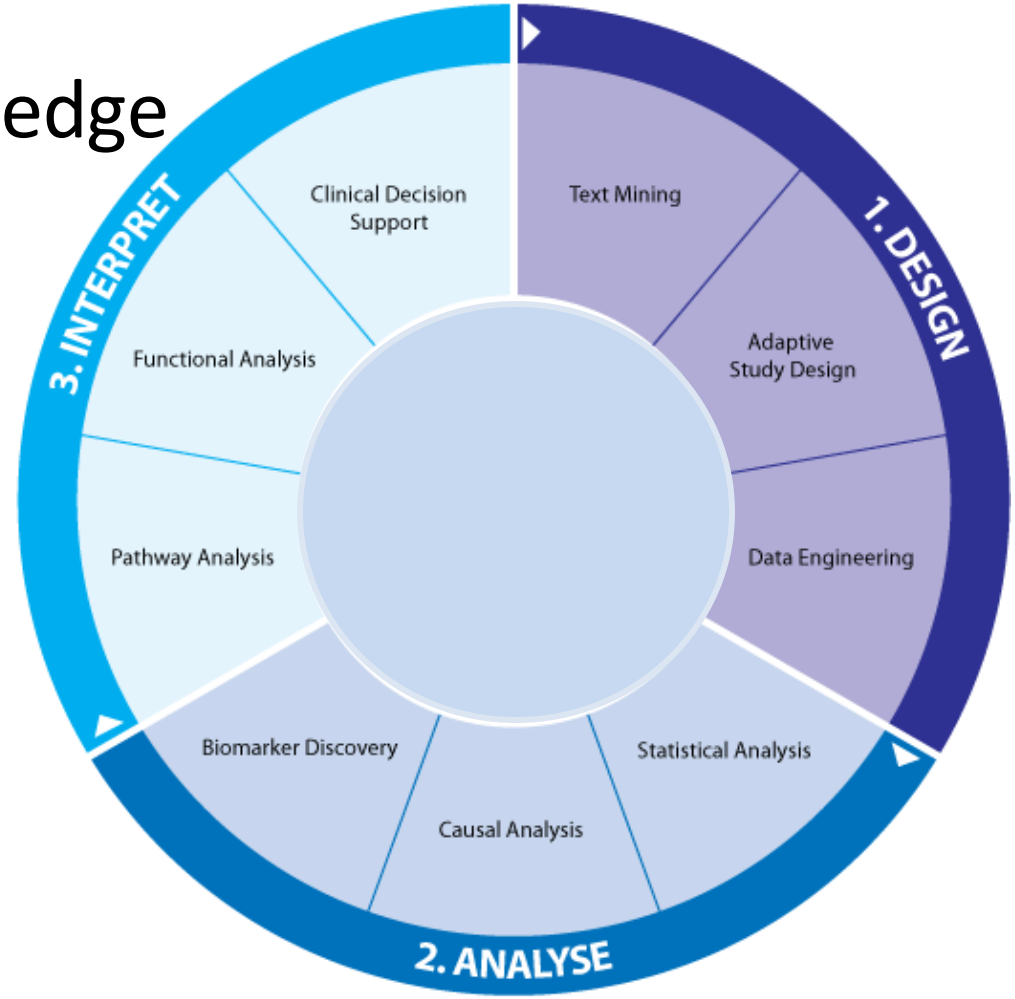
Special domains with workgroup option

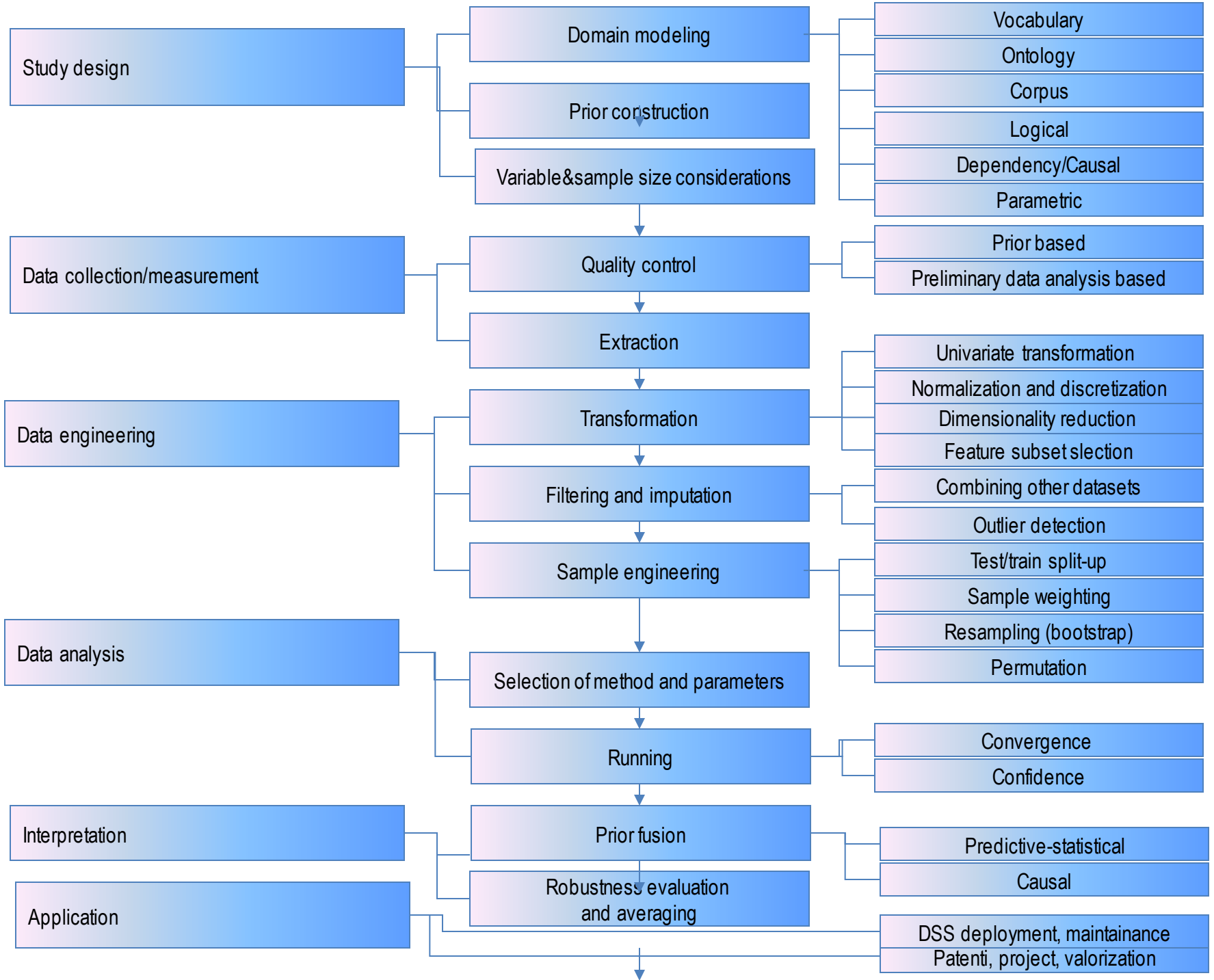
- **COVID-19**
- **Distance education**

- **Workgroup option**
  - **7+/-2 participants**
  - **complete intelligent data analysis study**
    - **Data collection.. model deployment**

(Biomedical) Data analysis in practice

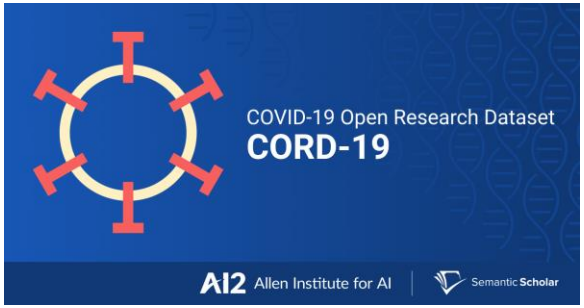
- Text mining/knowledge engineering
- Study design
- Data engineering
- Data analysis
- Interpretation
- Application




Study design
Domain modeling
Prior construction
Variable&sample size considerations
Vocabulary
Ontology
Corpus
Logical
Dependency/Causal
Parametric
Data collection/measurement
Quality control
Extraction
Prior based
Preliminary data analysis based
Data engineering
Transformation
Filtering and imputation
Sample engineering
Univariate transformation
Normalization and discretization
Dimensionality reduction
Feature subset slection
Combining other datasets
Outlier detection
Test/train split-up
Sample weighting
Resampling (bootstrap)
Permutation
Data analysis
Selection of method and parameters
Running
Convergence
Confidence
Interpretation
Prior fusion
Robustness evaluation and averaging
Predictive-statistical
Causal
Application
DSS deployment, maintainance
Patenti, project, valorization

COVID-19

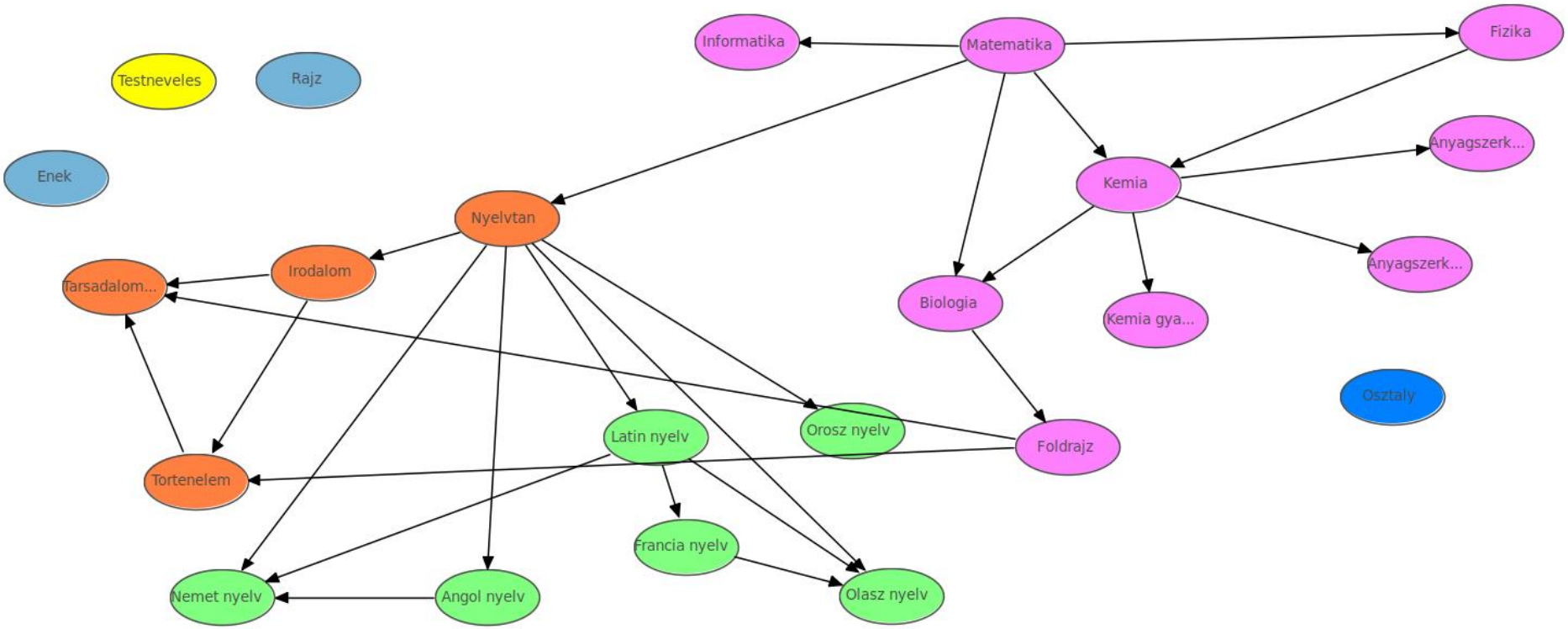
- Extremely varying symptoms/outcomes
  - personalized medicine
- Main medical goal:
  - early prediction of high risk patients
- Subgoals
  - Differential diagnosis
  - Multimorbidities
  - Adverse drug interactions
- Kaggle COVID-19 Open Research DatasetChallenge (CORD-19)



Distance education

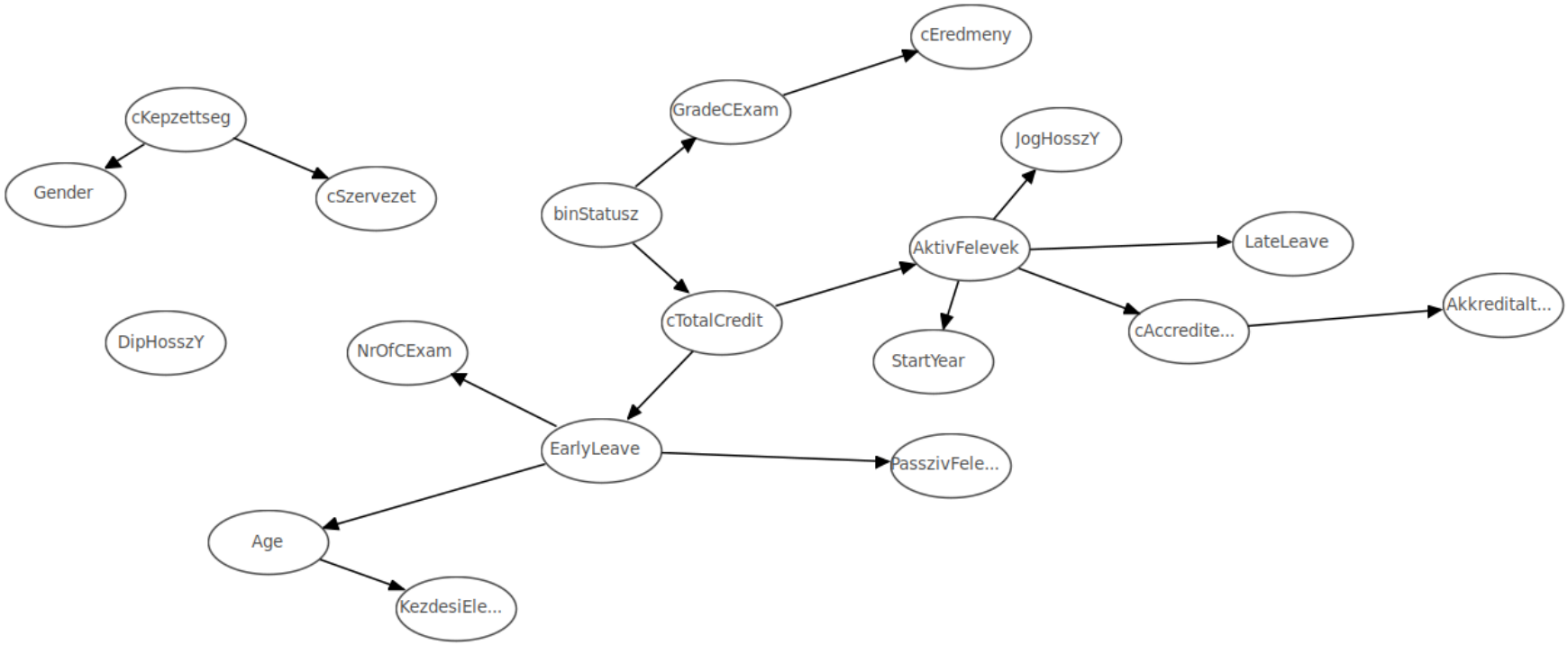
- Earlier Bayesian network models in education
  - Grade prediction in gymnasium
  - Attrition modeling in BME MSc
  - Student performance in artificial intelligence

Grade prediction



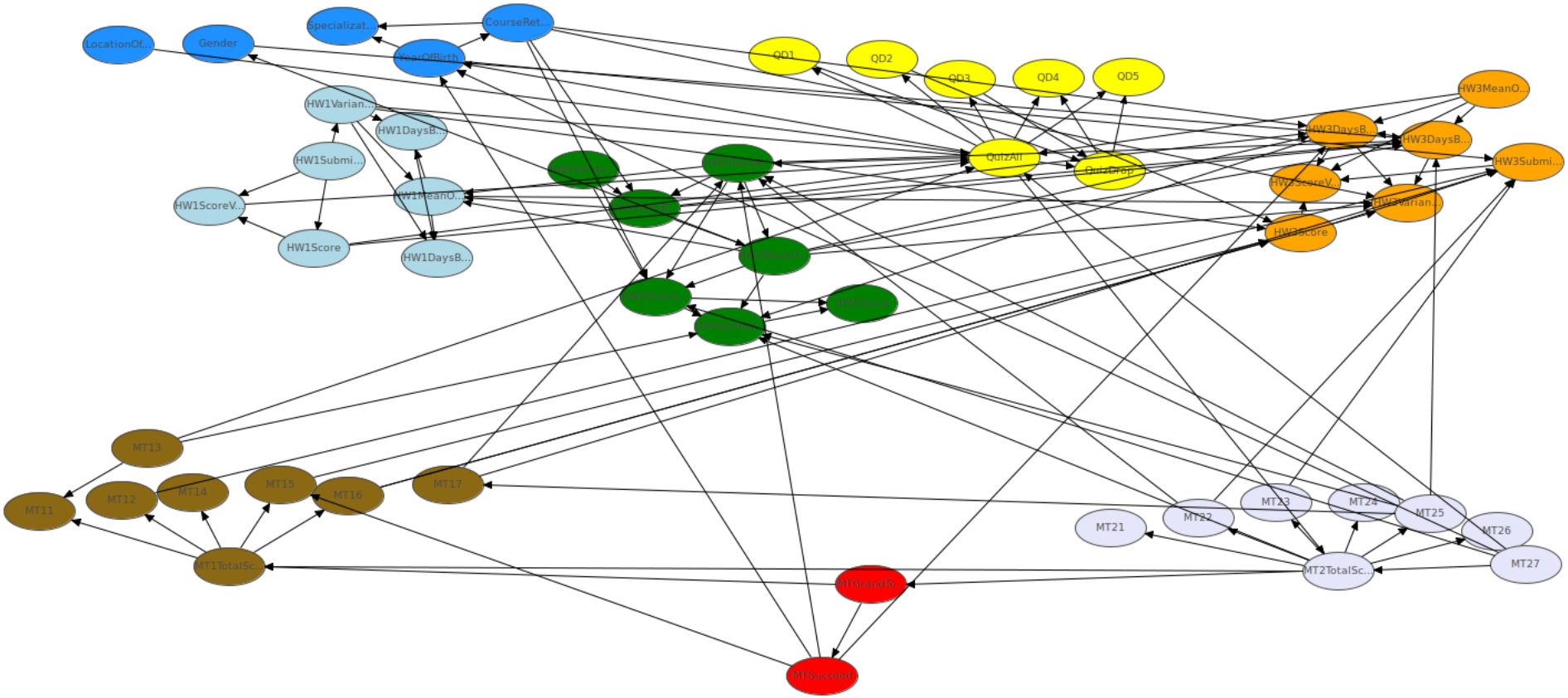
A.I.: BN homework guide

Student attrition



A.I.: BN homework guide

Student performance in AI



A.I.: BN homework guide

Distance education

- Side-effect and interplay with COVID-19
  - Quarantine, lockdown, emergency state
  - strategies
- Effect on performance and quality of life
  - academic performance
  - mental health
  - physical health
  - global happiness
- ++ massive open online courses (MOOCs)

Full fledged decision support

- Text mining/knowledge engineering
- Study design
- Data engineering
- Data analysis
- Interpretation
- Application

Homework

- Tools
  - BayesCube
    - Manual
    - API
  - Pomegranate

The **default** part

- **Default part:**
  1. Select a domain and sketch the structure of a Bayesian network model.
  2. Quantify your BN model.
  3. Check it with global inference and „information sensitivity of inference” analysis.
  4. Check it by relearning it from self-generated data.
  5. Demonstrate observational, causal, and counterfactual inference in the model.
  6. Extend your BN model to a decision network.
  7. Investigate the value of further information.

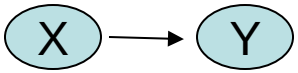
Homework steps: drafting

- **Default part:**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Consult it.

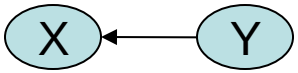
Direct?

Causal?

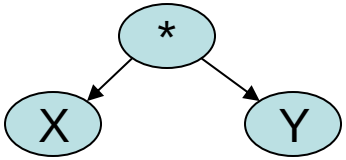
Causal models:



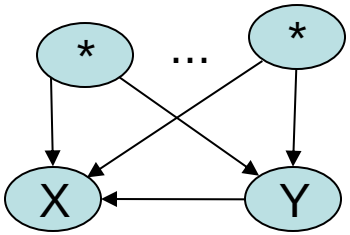
X causes Y



Y causes X



...



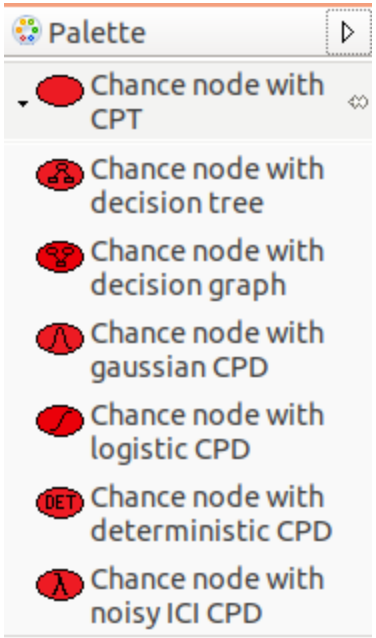
There is a common cause
(pure confounding)

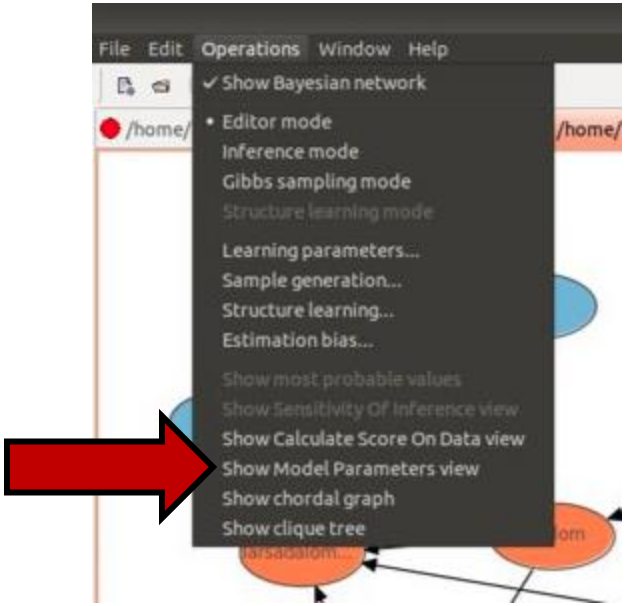
Consultation

The preliminary approval of your planned homework is mandatory!

Quantification: canonical models

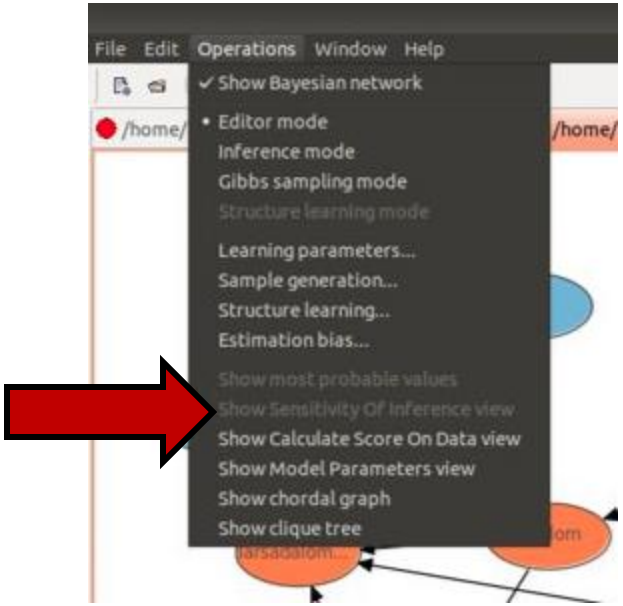
- **Default part:**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Consult it.
  - Quantify your BN model.

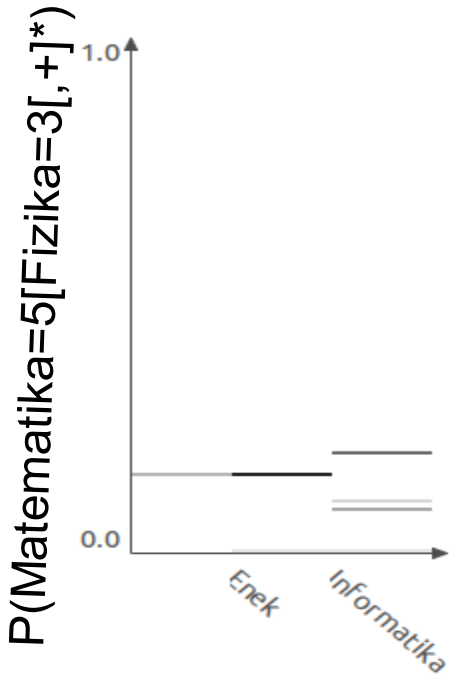


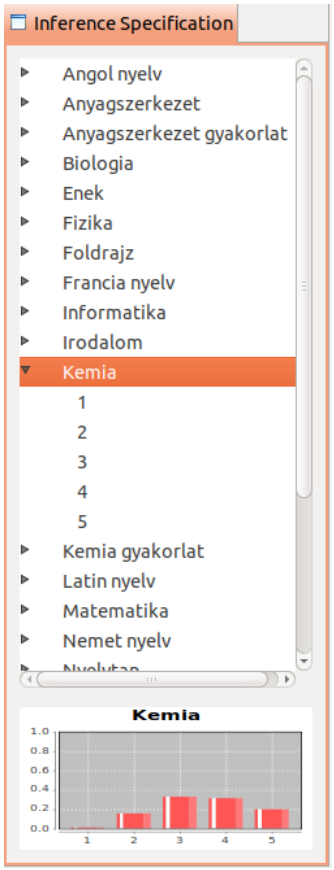


Check by test cases

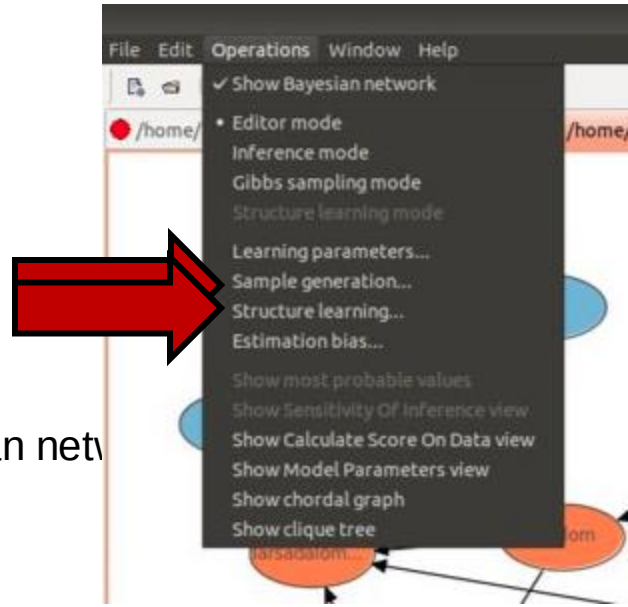
- **Default part:**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Consult it.
  - Quantify your BN model.
  - Check it with global inference
    and „information sensitivity of inference” analysis.



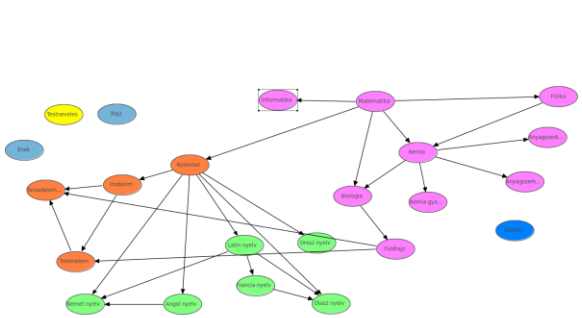


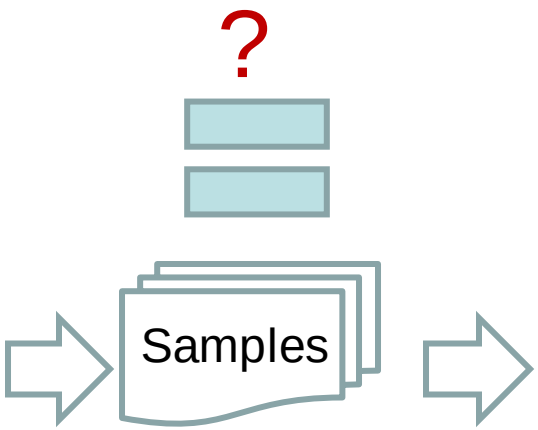


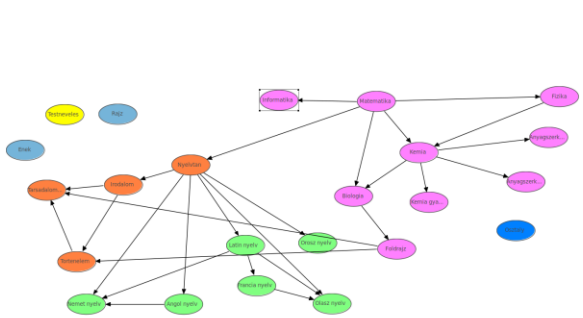
Global sanity check



- **Default part:**
  - Select a domain and sketch the structure of a Bayesian net
  - Consult it.
  - Quantify your BN model.
  - Check it with global inference and „information sensitivity of inference” analysis.
  - Check it by relearning it from self-generated data.
    - Generate a data set from your model.
    - Learn a model from your data.
    - Compare the structural and parametric differences between the two models.

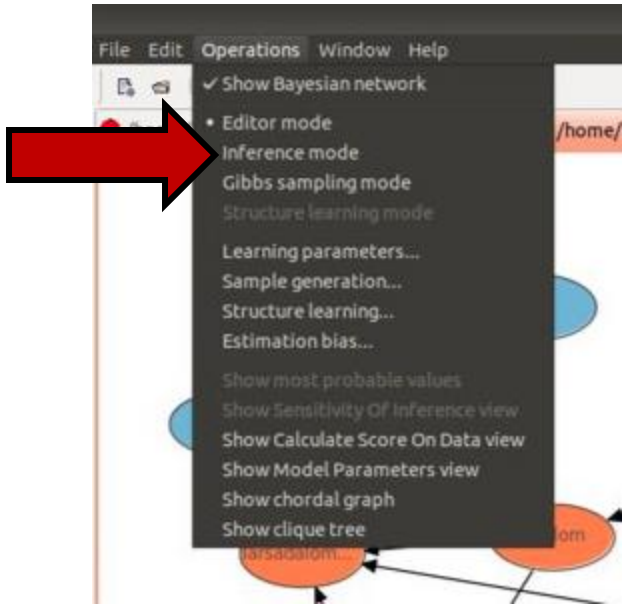






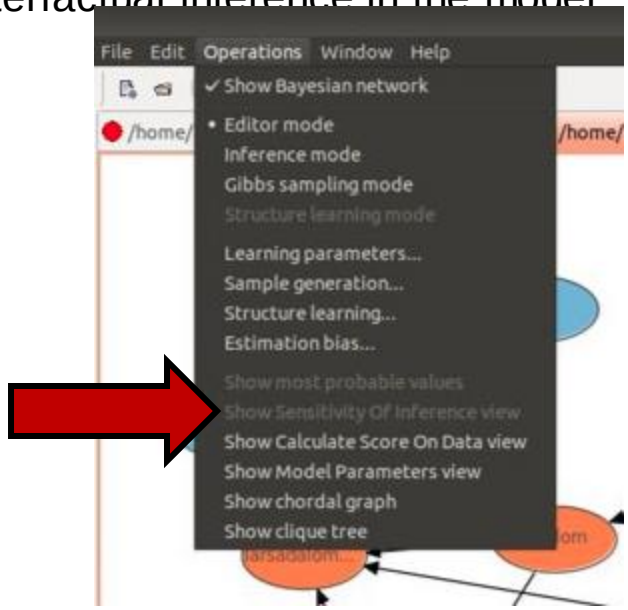
Demo: observational, causal, counterfactual inference

- **Default part:**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Consult it.
  - Quantify your BN model.
  - Check it with global inference and „information sensitivity of inference” analysis.
  - Check it by relearning it from self-generated data.
  - Demonstrate observational, causal, and counterfactual inference in the model.



Decision support

- **Default part:**
  - Select a domain and sketch the structure of a Bayesian network model.
  - Consult it.
  - Quantify your BN model.
  - Check it with global inference and „information sensitivity of inference” analysis.
  - Check it by relearning it from self-generated data.
  - Demonstrate observational, causal, and counterfactual inference in the model.
  - Extend your BN model to a decision network.
  - Investigate the value of further information.



Subtask: test a decision network

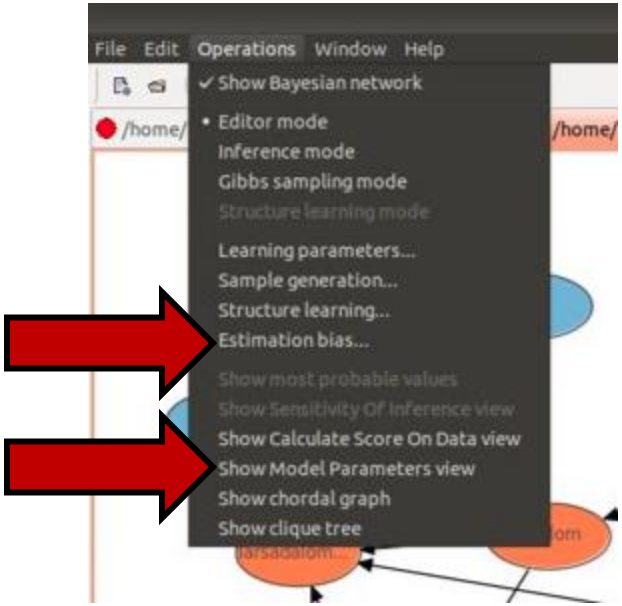
- Investigate the value of further information as follows:
  - select values for some “evidence” variables (E=e),
  - using BayesCube calculate the current expected loss/utility EU(D|e),
  - select a variable “I” as potential “further” information,
  - using BayesCube calculate the conditional probabilities of potential further observations (i.e. the conditional probabilities of potential values of this “further information” variable, p(I=i|E=e)),
  - using BayesCube calculate the expected losses/utilities corresponding to these potential further observations EU(D|e,i),
  - calculate the (expected) value of (perfect) information corresponding to this variable “I”, $\Sigma i\ p(i|e)*EU(D|e,i)- EU(D|e)$.

The "midterm" part

- Software environment
  - Pomegranate (to explore other environments, see the MI Almanach option)
- Expected format
  - Notebook (Google colab or Azure notebook)
  - Extended structure of the pomegranate colab
- Steps (~sections in a notebook):
  - Write a formal specification for your model with test cases.
  - Perform ALL(!) the tasks in the default homework
    - Construct a causal Bayesian network.
    - Test it by inference, sensitivity/perturbation/bootstrap analysis.
    - Demonstrate observational, causal, and counterfactual inference.
    - Extend into a decision network and infer optimal actions.
  - Perform and document additional steps either in BayesCube or pomegranate
    - Analyse estimation biases
    - Investigate the effect of model uncertainty and sample size on learning.

Midterm HW: estimation bias

- Midterm tasks
  - Write a formal specification for your model with test cases.
  - Construct a model using pomegranate or BayesCube AP
  - **Analyse estimation biases**.



Midterm HW: effect of model uncertainty and sample size on learning

- Midterm tasks
  - Analyse estimation biases.
  - **Investigate the effect of model uncertainty and sample size on learning**: vary the strength of dependency in the model (increase underconfidence to decrease information content) and sample size and see their effect on learning.

Scoring

Each subtask will get a mark and their average will be used to compute the final grade.

AI Almanach option for midterm

- Mesterséges Intelligencia (MI) Almanach
  - http://mialmanach.mit.bme.hu/
- Summary + demo of a software for PDSS
  - for PGMs/BNs/Bayesian inference/probabilistic programming
  - Suggested environments
    - pyBBN, pgmpy, libpgm, PyMC3, BUGS, PRISM, Stan
  - Suggested domain
    - the ADAS example
- Expected format
  - Notebook (Google colab or Azure notebook)
  - Structure of the pomegranate colab

Summary

- Select a domain, create variables (5-10), and specify structure.
- Quantify the Bayesian network.
- Analyse estimation biases
- Evaluate it with „information sensitivity of inference” analysis.
- Perform causal and counterfactual inferences.
- Generate a data set from your model.
- Learn a model from your data.
- Compare the structural and parametric differences between the two models.
- Evaluate value of further information.
- Investigate the effect of model uncertainty and sample size on learning.