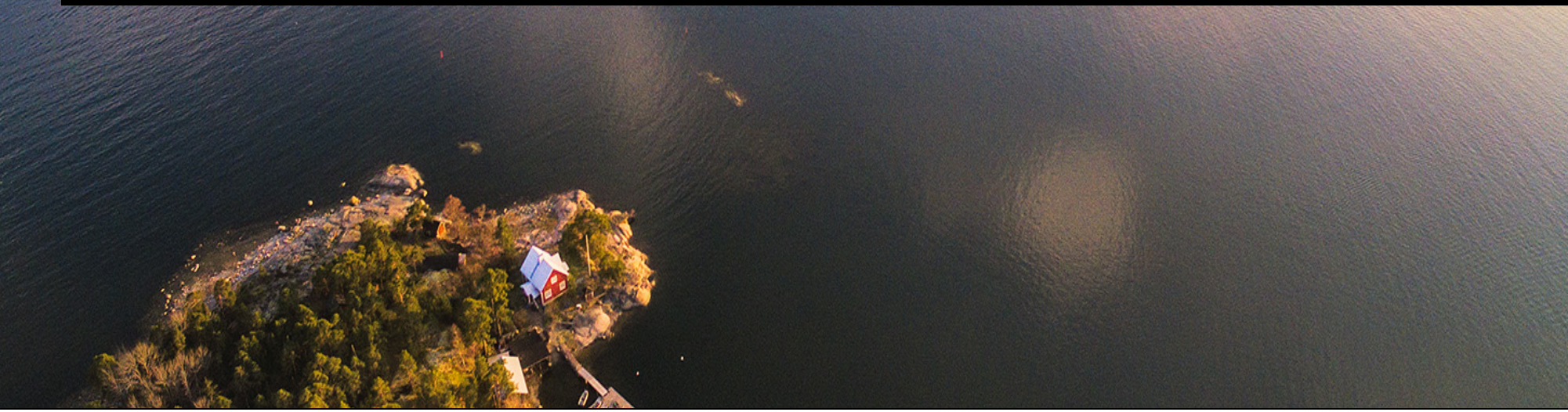
Structural elicitation for Bayesian Networks

Joint work with Anca Hanea (CEBRA, UniMelb) and Sophia Wright (Univ of Warwick)

Tina Nane (TU Delft)



TU Delft

Citation Performance of Researchers

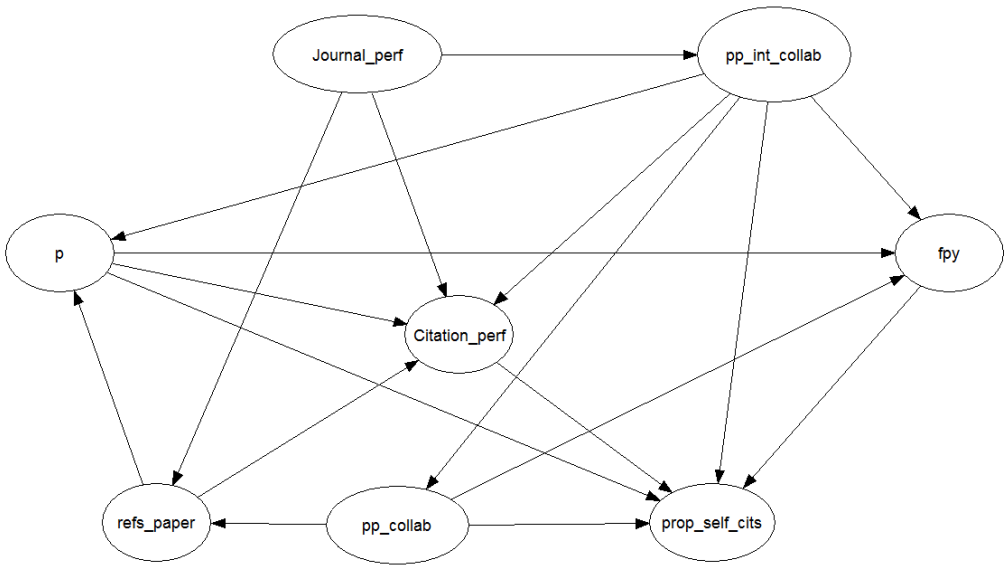
Influencing factors

- Publication record, years of activity
- Journal citation scores
- Field
- (International) collaboration, etc.

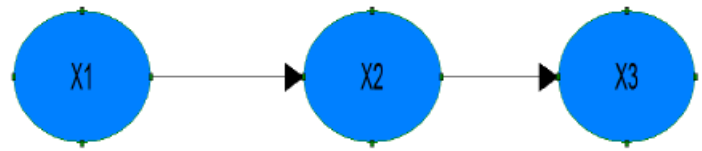
Citation Performance of Researchers

Influencing factors

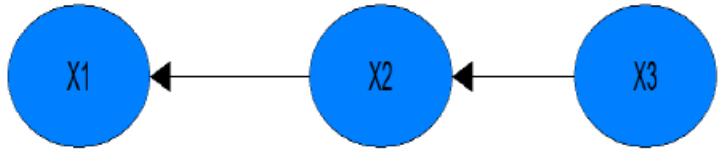
- Publication record, years of activity
- Journal citation scores
- Field
- (International) collaboration, etc.



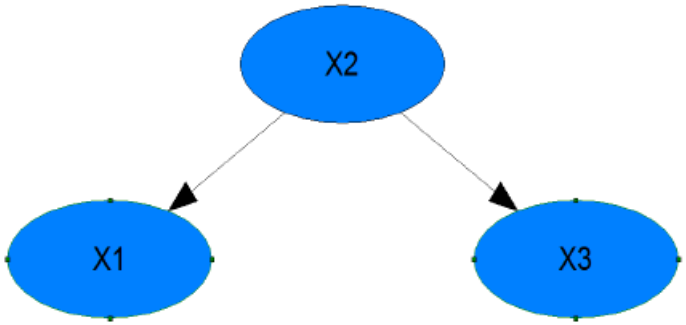
Bayesian Networks



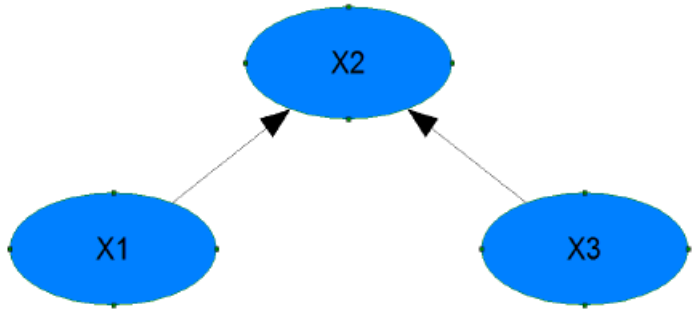
a)



b)

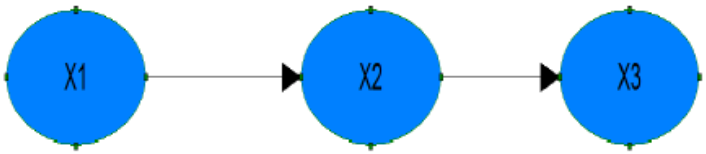


c)

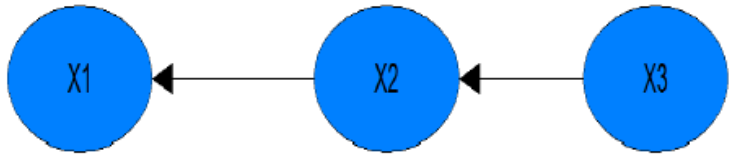


d)

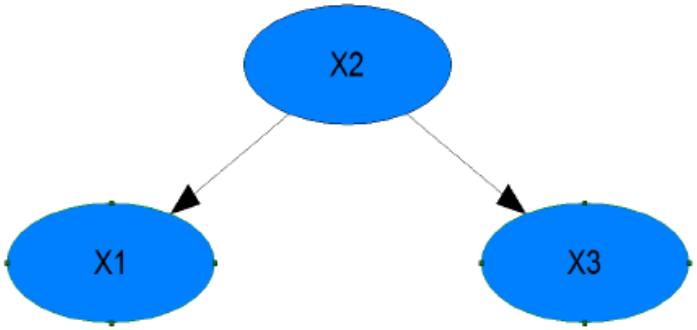
Bayesian Networks



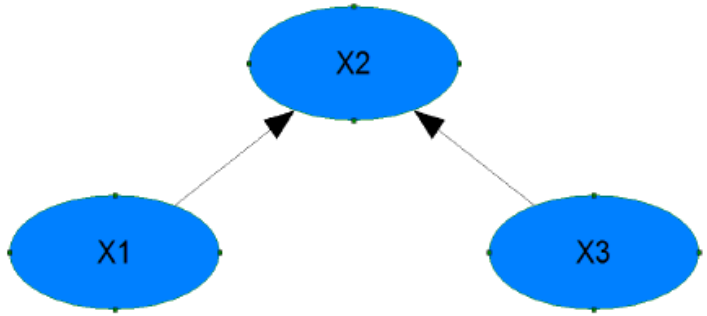
a) $X_1 \perp X_3 | X_2$



b) $X_1 \perp X_3 | X_2$



c) $X_1 \perp X_3 | X_2$



d) $X_1 \not\perp X_3 | X_2,\ X_1 \perp X_3$

Bayesian Networks

Number of possible structures

n	2	3	4	5	10
nr of DAGs	3	25	543	29281	4.2×10^{18}

Bayesian Networks

Number of possible structures

n	2	3	4	5	10
nr of DAGs	3	25	543	29281	4.2×10^{18}

Learning the structure of a Bayesian Network

- Experts
  - Write all the variables of interest
  - Write all variables that could influence the variables of interest
  - Write parents of these variables, etc.
- Data driven
  - Constraint based algorithms
  - Score based algorithms
- Experts + Data

Learning the structure of a BN

- Need for a performance-based elicitation protocol
- How can we measure performance when eliciting the structure of a Bayesian Network?

Learning the structure of a BN

- Need for a performance-based elicitation protocol
- How can we measure performance when eliciting the structure of a Bayesian Network?

Our approach

- Expert
  1. Ask experts about the conditional distribution of the variable of interest

Learning the structure of a BN

- Need for a performance-based elicitation protocol
- How can we measure performance when eliciting the structure of a Bayesian Network?

Our approach

- Expert
  1. Ask experts about the conditional distribution of the variable of interest
- Data
  2. Assign arcs in particular order
  3. Compute the conditional distribution of the variable of interest
  4. Repeat 2 & 3

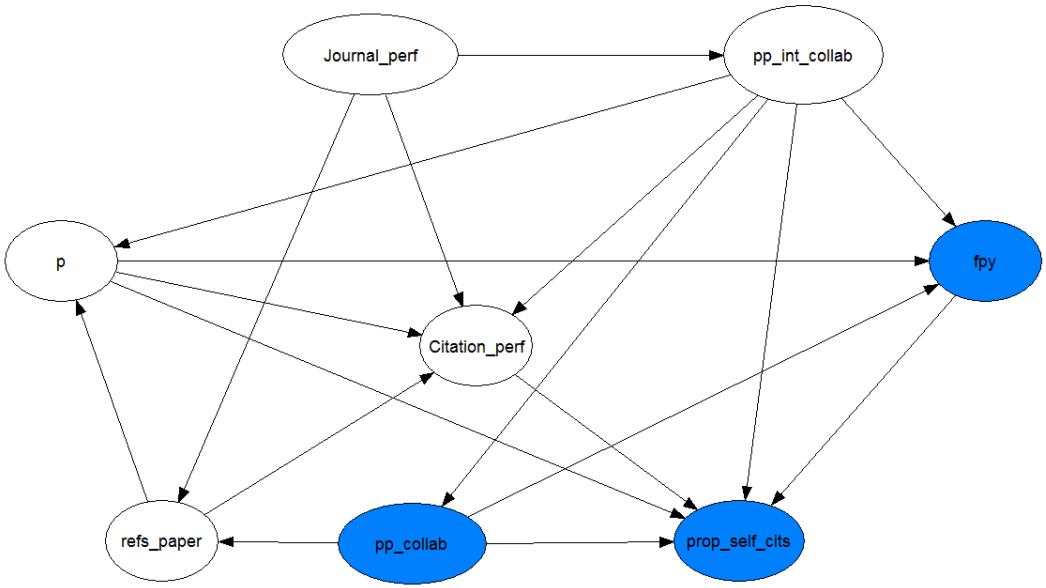
Learning the structure of a BN

- Need for a performance-based elicitation protocol
- How can we measure performance when eliciting the structure of a Bayesian Network?

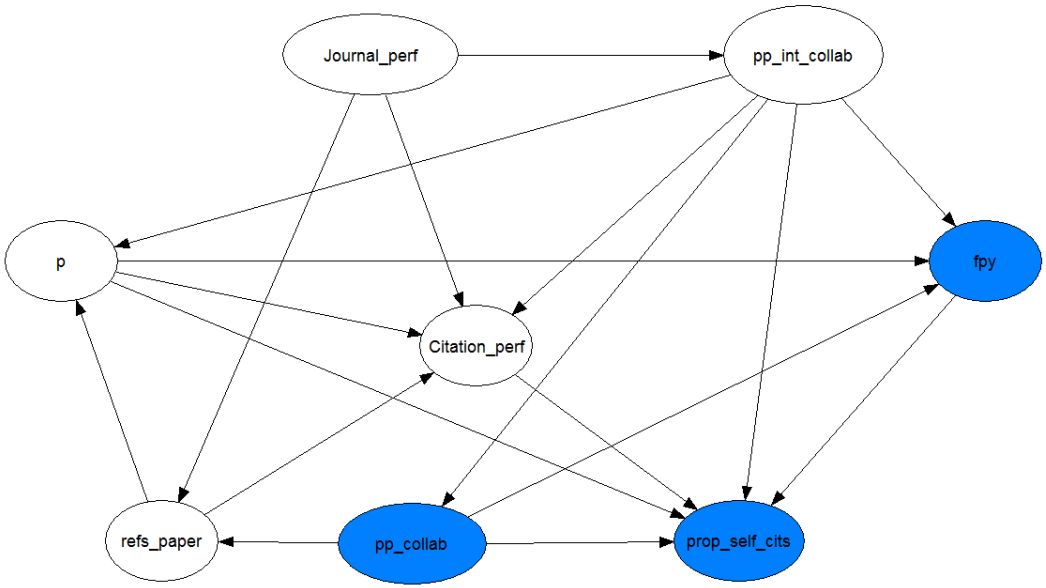
Our approach

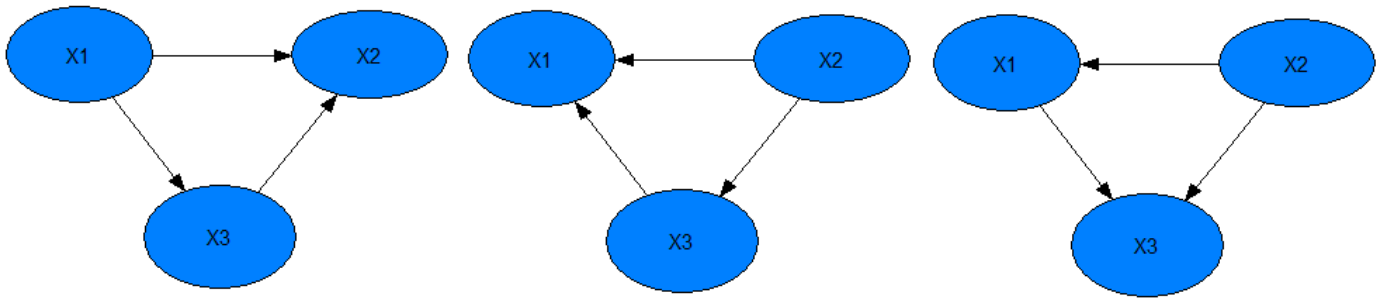
- Expert
  1. Ask experts about the conditional distribution of the variable of interest
- Data
  2. Assign arcs in particular order
  3. Compute the conditional distribution of the variable of interest
  4. Repeat 2 & 3
- Compare the conditional distributions in 1 and 3
- Choose the conditional distribution from data closest to the conditional distribution from experts (with respect to a particular distance)

Citation performance of researchers

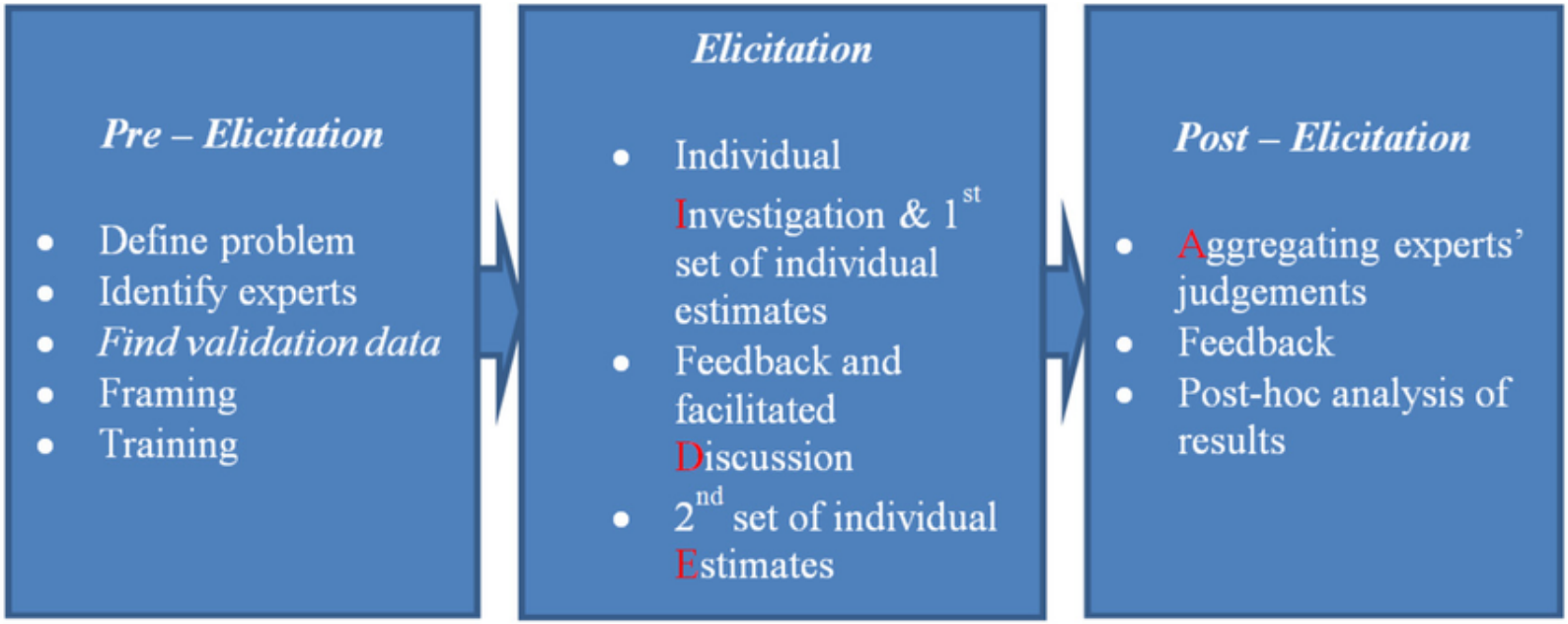


Citation performance of researchers

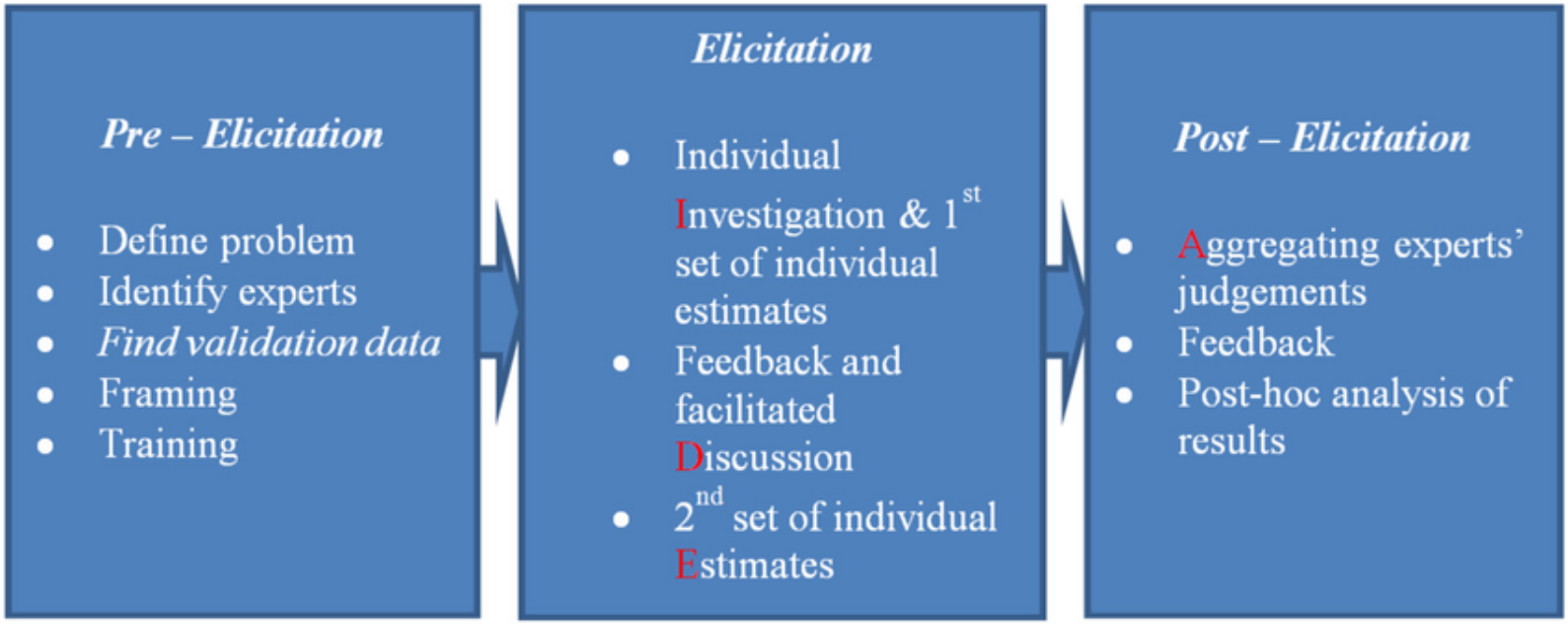




IDEA protocol



IDEA protocol



Two rounds of the Classical Model, intermediated by feedback and facilitated discussion

Elicitation results

Round one

Id	Calibr.	Mean relative realization	Normaliz.weight without DM
A	0.01397	1.183	0.0553
B	0.2895	0.5229	0.5067
C	0.06083	0.6187	0.126
D	0.06372	0.7125	0.152
E	0.2895	0.1651	0.16
DM_{perf}	0.4735	0.1377	

Elicitation results

Round one

Id	Calibr.	Mean relative realization	Normaliz.weight without DM
A	0.01397	1.183	0.0553
B	0.2895	0.5229	0.5067
C	0.06083	0.6187	0.126
D	0.06372	0.7125	0.152
E	0.2895	0.1651	0.16
DM_{perf}	0.4735	0.1377	

Round two

Id	Calibr.	Mean relative realization	Normalized weight without DM
A	0.04706	0.7362	0.06141
B	0.4735	0.3703	0.3108
C	0.6827	0.3883	0.4699
D	0.06372	0.5211	0.05885
E	0.2895	0.1931	0.09908
DM_{perf}	0.6827	0.1418	

Elicitation results

Influencing factors

EXPERT A	EXPERT B	EXPERT C	EXPERT D	EXPERT E
1. Journal score	1. Journal score	1. pp_int_collab	1. Journal score	1. refs_paper
2. pp_int_collab	2. pp_int_collab	2. Journal score	2. pp_int_collab	2. Journal score
3. Output	3. pp_collab	3. pp_collab	3. pp_collab	3. pp_int_collab

Elicitation results

Influencing factors

EXPERT A	EXPERT B	EXPERT C	EXPERT D	EXPERT E
1. Journal score	1. Journal score	1. pp_int_collab	1. Journal score	1. refs_paper
2. pp_int_collab	2. pp_int_collab	2. Journal score	2. pp_int_collab	2. Journal score
3. Output	3. pp_collab	3. pp_collab	3. pp_collab	3. pp_int_collab

Conditioning

- Given that the Journal score is at its 95% quantile and the International collaboration score is at its 95% quantile and the Output is at its 95% quantile, what are your estimates for the average citation performance of a researcher?

Elicitation results

Influencing factors

EXPERT A	EXPERT B	EXPERT C	EXPERT D	EXPERT E
1. Journal score	1. Journal score	1. pp_int_collab	1. Journal score	1. refs_paper
2. pp_int_collab	2. pp_int_collab	2. Journal score	2. pp_int_collab	2. Journal score
3. Output	3. pp_collab	3. pp_collab	3. pp_collab	3. pp_int_collab

Conditioning

- Given that the Journal score is at its 95% quantile and the International collaboration score is at its 95% quantile and the Output is at its 95% quantile, what are your estimates for the average citation performance of a researcher?

Results

Percentile		5%	50%	95%
Round	1.	15.88	28.19	51.91
	2.	8.531	30.61	48.81
Data		7.5	31.1	57.143

Discussion and conclusions

- Regardless the arcs assignment, the conditional distribution of the variable of interest might not change
- Given a particular structure of the BN, the experts can assess conditional distributions of the variable of interest quite accurately
- IDEA can help experts to become more calibrated, with a possible small decrease in information score
- IDEA can increase the performance of the DM

Thank you!