Structure Elicitation Approaches for Bayesian Networks: A Survey

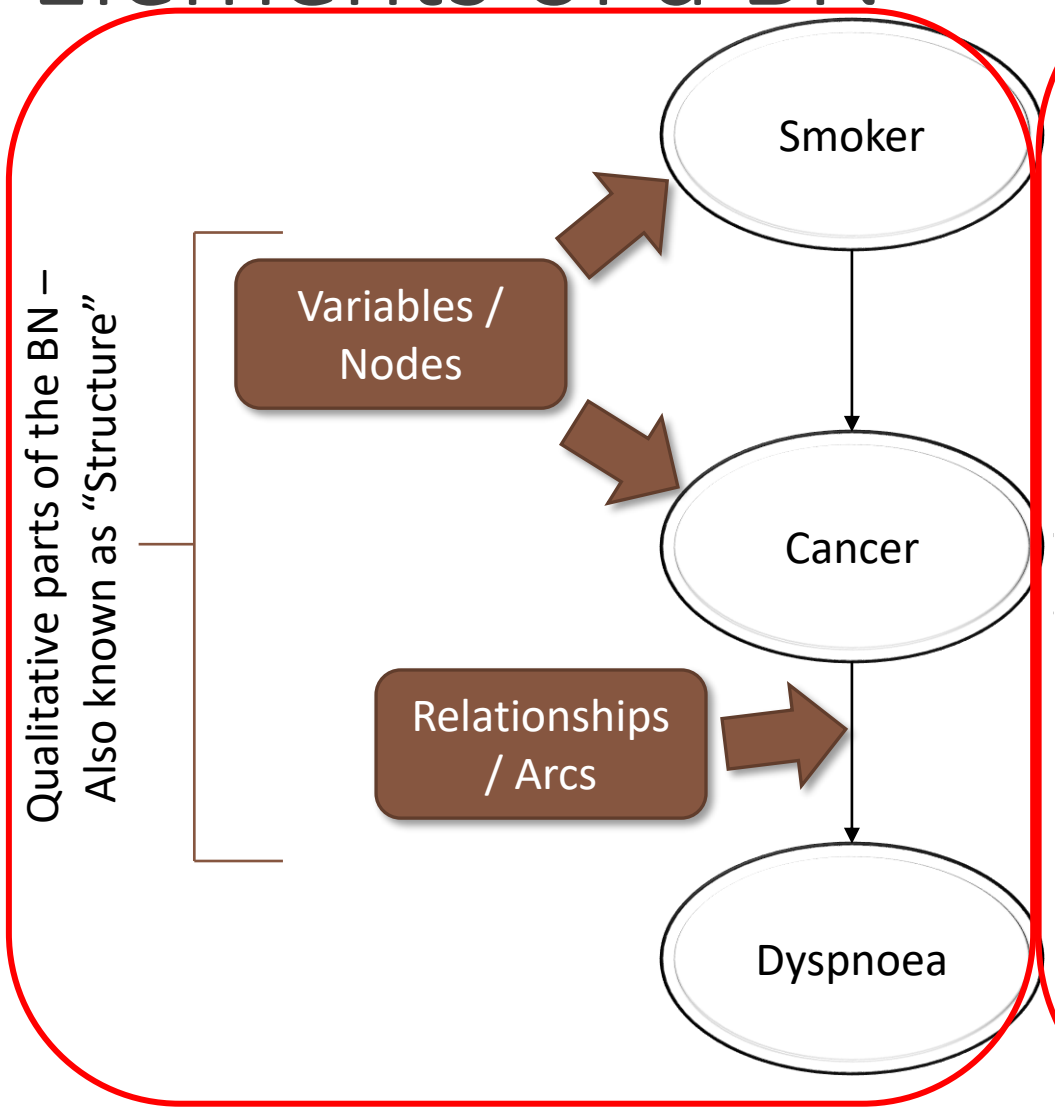
ROSS PEARSON

DR. STEVEN MASCARO

PROF. ANN NICHOLSON

MONASH University

Elements of a BN



Smoker = True	Smoker = False
0.2	0.8

Parameters

Quantitative part of the BN

	Cancer = True	Cancer = False
Smoker = True	0.1	0.9
Smoker = False	0.02	0.98

	Dyspnoea = True	Dyspnoea = False
Cancer = True	0.45	0.55
Cancer = False	0.01	0.99

Why apply elicitation during BN construction?

- Data unavailable
- Expert knowledge provides a better causal explanation
- Model validation
- Priors for structure learning

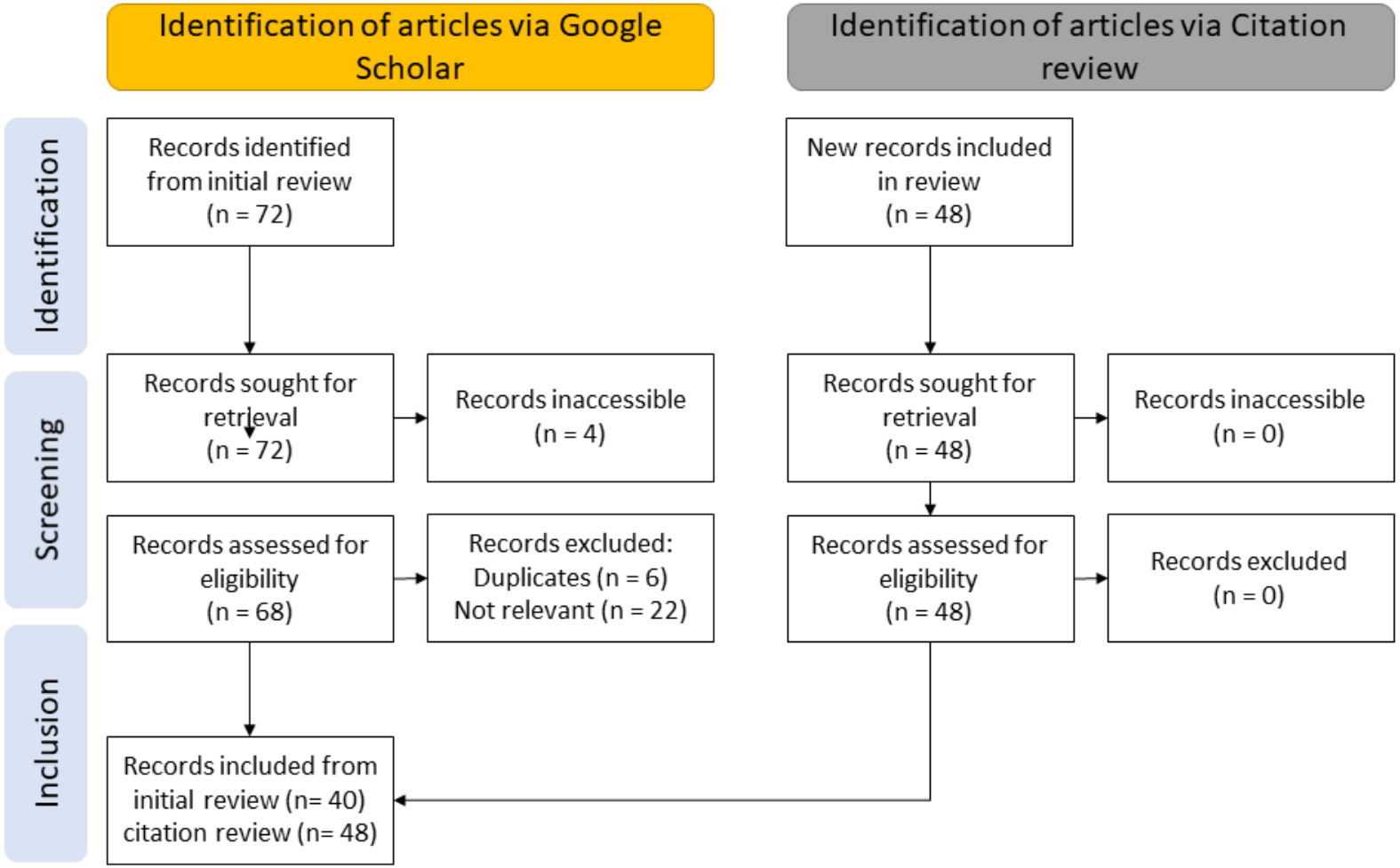
Elicitation of BN Structure

- Little attention within literature.
- Structure elicitation = an amateurish art (Burgman et al., 2021)

Motivation

Advancing approaches to structure elicitation (and elicitation in general) is to address the expert bottleneck (Ratnapinda & Druzdzel, 2011).

Semi Structured Survey Methodology



Steps of Structure Elicitation

1. Background and preparation,
2. Recruiting and training expert(s),
3. Structure Elicitation, and
4. Post elicitation

Step 1 - Background and preparation

- Problem identification
  - Natural language description of the problem
  - Model purpose & underlying assumptions
  - Is a BN even appropriate?
- Existing models
  - Converting a causal map into a BN (Nadkarni & Shenoy, 2004)
  - Building a BN from an ontology (Fenz, 2012; Helsper, 2001)
  - Influence diagrams (Chen & Pollino, 2012)

Elicitation considerations (Step 1)

- Elicitation Challenges
  - Elicitation burden (time and resources required)
  - Cognitive biases and heuristics
  - Different forms of uncertainty - Aleatory and epistemic.
- Knowledge Reuse
  - Templates (Rajabally et al., 2004),
  - Object oriented BNs (Koller & Pfeffer, 1997),
  - Network fragments (Laskey & Mahoney, 1998),
  - Idioms (Neil et al., 2000), and
  - Ranked nodes (N. E. Fenton et al., 2007).

Structure Elicitation Tools (Step 1)

- BN modelling tools: Hugin, Netica, AgenaRisk, Genie, etc.
- OTS general software: Google sheets and Microsoft Excel, etc.
- Established cloud based: Qualtrics
- Custom BN “structure” elicitation software:
  - The BARD platform (Bayesian Argumentation via Delphi) (Nyberg et al., 2021)
  - LinkIT and SortIT (Cao, 2014)
  - Bayesian Network Elicitor (BNE) (Serwylo, 2016)
  - Interlink JPathfinder tool (Rizzo & Blackburn, 2018).

Step 2 - Recruiting and training experts

- Degree of expertise of experts (Dunn, 2016).
- Sample variable sets/structures/models (Baker, 2009)
- Calibration of experts; and
- Feedback (Beaudrie et al., 2016; O'Hagan et al., 2006; Werner et al., 2017)
- Preparatory steps (Nicholson et al., 2016)

Roles during elicitation (Step 2)

- Decision maker: customer or consumer
- Substantive expert: subject matter expert or expert
- Statistician: BN expert or engineer or modeller
- Facilitator

Step 3 - BN Structure Elicitation

- Iterative
- Individually e.g. (Flores et al., 2011) or in groups,
- Face-to-face or via video conferencing or some other digital interface e.g. BARD (Nyberg et al., 2021),
- Through survey or interview e.g. (Arentze et al., 2008),
- Locally or remotely e.g. (E. R. Pearson et al., 2021).
- Self elicitation (O'Hagan et al., 2006).

Variables (Step 3)

- Variable identification:
  - literature,
  - data or
  - during elicitation
  - Existing models
- Ranking variables

Relationship types (Step 3)

Relationship	Description	Example
Causal	Where two variables share a direct and directional relationship and this relationship is understood to be causal.	$A \rightarrow B$
Direct	Where two variables share a direct and directional relationship and this relationship is understood to be non causal.	$A \rightarrow B$
Correlation	Where two variables share a relationship, it is not known if this relationship is direct or otherwise, it is also not known which variable causes which. A correlation simply indicates that when either variable changes the other will also change.	$A \sim B$
Pairwise non-directional	Where two variables share a direct relationship but it is unknown which variable causes which.	$A - B$
Indirect / Causal dependency	Where two variables share a relationship indirectly. That is there are intermediate variables and direct relationships that describe this indirect relationship. An indirect relationship differs from correlation in that a) the direction is known, and b) the relationship is known not to be direct.	$A \Rightarrow B$
Independant	Where two variables are known not to have a direct relationship and that they are independent from one another (unconditional independence).	$A \not\Rightarrow B$
Temporal tier	Where variables conceptually exist within temporal tiers. Relationships between a variable in a subsequent tier and a variable in a prior tier is not permissible.	$A \prec B$

Adjacency Matrices (Step 3)

	Cancer	Pollution	Smoker	X-ray	Dyspnoea
Cancer				→	→
Pollution	→				
Smoker	→				
X-ray					
Dyspnoea					

Tools and Methods for Structure Elicitation (Step 3)

- Pairwise relationships
- Questionnaires
- Interviews
- Surveys
- Group Interviews
- Temporal Tiers
- Causal Maps

Step 4 - Post Structure Elicitation – Structure Aggregation

- Challenges:
  - variable incongruity,
  - avoiding cycles,
  - preserving conditional independencies,
  - resolving ambiguous arc directions

- Approaches
  - Flores et al. combined multiple BNs into a single aggregate, by ordering each arc based on the frequency it appeared.
  - Serwyo (Serwylo, 2016) considered two other approaches for structure aggregation, namely majority vote and the expectation maximum (EM) algorithm (Dempster et al., 1977).
  - Topological fusion, a process described by (Matzkevich & Abramson, 1992)
  - Causal Structure Aggregation Model (CSAM) (Baker, 2009)

Step 4 - Post Structure Elicitation – Structure Evaluation

- Is the model a faithful representation of the problem (Face Validation)
- Sensitivity analysis -sensitivity to findings and sensitive to parameters
- Structure learning can be used to validate elicited structure (or vice versa)
- Gold Standard
- Edit distance and Structural Hamming Distance (SHD)
- Difference matrices

Final Thoughts

- Systematic processes have been demonstrated to improve elicitation results
- There is a need for structure elicitation approaches despite a lack attention in the literature
- Existing approaches for general elicitation and parameter elicitation can be draw from as useful references
- Structure elicitation can be useful where: you have no data, you want to elicit structure priors for structure learning, expert knowledge is desired for construction of the qualitative part of a BN, or you want another BN structure to support validation activities.

Next Steps…

- Development of a protocol for structure elicitation