Customising a Graph to Support a Structural Elicitation

Jim Q. Smith

University of Warwick and Alan Turing Institute

July 2017 Delft

Themes & Content of this talk

- **Structural information** is often more **faithful, robust & universal** than quantitative.
- Essential in **multivariate contexts** plus when supporting **many agents** & striving to **draw these together**
- **First** elicit structural beliefs & **only then** quantitative embellishments.
- **Use available data** to focus & support embellishment & critique structure.
- Structural information is often well expressed through **graphs**.
- But argue here essential to **customise the graph** to each elicited description of a domain.
- Will illustrate **BN methods translate** to other graphical structural hypotheses.
- Illustate how most ubiquitous & robust knowledge - especially within decision support - expressed through **casual hypotheses** within these customised domains.

Structural expression through a BN: Barclay et al (2013)

$$\begin{tikzcd}[column sep=small]
 & & \text{Social Background } s \arrow[dll] \arrow[drr] & & \\
\text{Economic Situation } e \arrow[rr] & & \text{Family Life Events } f \arrow[rr] & & \text{Hospital Admissions } h
\end{tikzcd}$$

$$p(s, e, f, h) = p(s)p(e|s)p(f|e)p(h|sf)$$

- Actually expresses just 2 qualitative statements

$$F \perp\!\!\!\perp S|E \quad \& \quad H \perp\!\!\!\perp E|S, F$$

- In expert's natural language: "Given a child's economic situation, occurrence of family life events does not depend on her social background" & "Once we know a child's social background & extent of family life events then the family's economic situation has no effect on whether or not that child is taken to hospital".

Bayesian estimation Diagnostics and Selection

- Priors on parameters $p(s), p(e|s), p(f|e).p(h|sf)$, Dirichlet (discrete) or Gaussian (continous)
- Observe n units respecting a BN on $p(s), p(e|s), p(f|e).p(h|sf)$,e.g multinomialor Gaussian.
- For each candiate M_i using form *marginal likelihood* $\pi_i(s, e, f, h)$

$$\pi_i(s, e, f, h) = \pi_i(s)\pi_i(e|s)\pi_i(f|e)\pi_i(h|sf)$$

- Diagnostics (e.g. prequential) then available for each model.
- Select model with highest score (MAP) on product evaluated at the observations.

Embedded Control & Causal BN's: Spirtes,Pearl,etc.

Causal hypotheses often linked to control. Arrows will often be read as:

$$\begin{tikzcd} & \text{Social Background } s \arrow[dl] \arrow[dr] & \\ \text{Economic Situation } e \arrow[r] & \text{Family Life Events } f \arrow[r] & \text{Hospital Admissions } h \end{tikzcd}$$

"Sort out a family's economic situation & its life events will improve to same extent regardless of its social status".
"If we can give the child a high social background & protect her from pernicious life events then money adds nothing more to her health".

- Ubiquitous big leap! Assumes structure of observed system right & controlling to state produces same results downstream as observing this state.
- If leap valid, then we have a powerful tool! Observational experiments link to controlled experiments and vice versa.

Comments about such BNs

Conditional independence can be:

- Stated directly in natural language by the expert (**faithful**). So $\mathcal{I}$ directly expresses this assertion. Also possible to check directly whether logical consequences (determined by semigraphoid axioms) sound reasonable when translated
- Often given a proverbial status and can be applied to a wide range of applications - not just current study. (**robust**)
- More easily agreed upon by others - straightforward to understand & perhaps as statements of received wisdom (**universal**)

THEREFORE

- **First** elicit expert's structural beliefs & **only then** its quantitative embellishments.
- Use available data to focus & support embellishment.
- Simple to extend to **causal** hypotheses e.g. "Sort out a family's economic situation and its life events will improve regardless of its social status".

BN's, Structure and Graphs

- Used in the right circumstances a BN can be a really useful tool to embed elicited & most secure structural information & then only quantitative information around this frame.
- Graphs provide a very accessible way of representing structure.
- Methods are especially helpful when domain information is causal.
- BNs are widely supported by software - some free - & so easily implemented.

HOWEVER

Many problems not naturally expressible a vanilla BNs. Despite availability of software - we shouldn't use them!!.

Question: Can we then construct different graphical methods enjoying advantages of BNs?

Answer: Yes we can!! Now many examples. Illustrate here some I have been personally developing.

A proposal for more general structural modelling

If a BN is not a natural way to expert of expressing her domain we shouldn't use it!!!! Instead:

- **Customise semantics** & associated **graph** to faithfully reflect types of relationships in given domain.
- Discover ways of directly **interrogating** graph using logical consequences of bespoke semantics.
- Develop **elicitation techniques** to **embellish** graph into **full probabilistic description** (while continuings to respect graph's assertions).
- Use probabilistic embellishments for **uncertainty handling** & for **embedding** any **data** to check efficacy of plausible structural hypotheses.
- For robustness try to use new semantics to find **mappings** from a graph of uncontrolled system to its controlled analogues.

Here are examples of a few recent graphical methods formally developed & successfully applied!!!

Different Dynamic Framework: Costa et al (2015)

- LMDM (Queen & Smith,1993) a graphical class of **dynamic** regression models.

$$
\begin{aligned}
Y_t(i)|PaY_t(i) &= \mathbf{F}_t(PaY_t(i))\boldsymbol{\theta}_t(i) + v_t(i) \\
\boldsymbol{\theta}_t(i) &= G\boldsymbol{\theta}_t(i) + \mathbf{w}_t(i)
\end{aligned}
$$

- Natural objects here **processes not variables.** Has BN representation but complicated: infinite # nodes. Even on $(Y_t(1), Y_t(2))$

$$
\begin{array}{ccccc}
\cdots & \boldsymbol{\theta}_t(1) & \rightarrow & \boldsymbol{\theta}_{t+1}(1) & \cdots \\
 & \downarrow & & \downarrow & \\
 & Y_t(1) & & Y_{t+1}(1) & \\
 & \downarrow & & \downarrow & \\
 & Y_t(2) & & Y_{t+1}(2) & \\
 & \uparrow & & \uparrow & \\
\cdots & \boldsymbol{\theta}_t(2) & \rightarrow & \boldsymbol{\theta}_{t+1}(2) & \cdots
\end{array}
$$

An Example on 4 processes

- Draw new graph to represent LMDM where nodes represent processes! Dependence on states kept implicit.

$$\begin{array}{ccccc} Y(1) & \rightarrow & Y(3) & \rightarrow & Y(4) \\ & \nearrow & & & \\ Y(2) & & & & \end{array}$$

- e.g. for component process $Y(3) = \{Y_y(3)\}$ read

$$\boldsymbol{\theta}_t(3) = (\theta_t(3,0), \theta_t(3,1), \theta_t(3,2))$$

$$\begin{aligned} Y_t(3) | Y_1(1), Y_1(2) &= \theta_t(0) + \theta_t(1) Y_1(1) + \theta_t(2) Y_t(2) + v_t(i) \\ \boldsymbol{\theta}_t(i) &= \boldsymbol{\theta}_t(i) + \mathbf{w}_t(i) \end{aligned}$$

- Note graph summarises transparently dependences - but **semantic quite different from BN**!!

Why a new graph and semantics useful

- Graph **depicts** hypothesised process unambiguously & simply.
- Embellishs into **full probabilistic model** with a few addtional elicitations (Just like for BN).
- Admits simple inference $\Rightarrow$ graphical model easily accomodates available data under conjugate closed form analysis. So e.g. BF scores are products of multivariate student t's: $\Rightarrow$ diagnostics can be built & e.g with many candidate models greedy search model selection over an apiroi elicited class of different plausible models!
- Unlike BN each graph depicts a different & distinguishable statistical processes."Causal" direction from independence of regression coeffs.

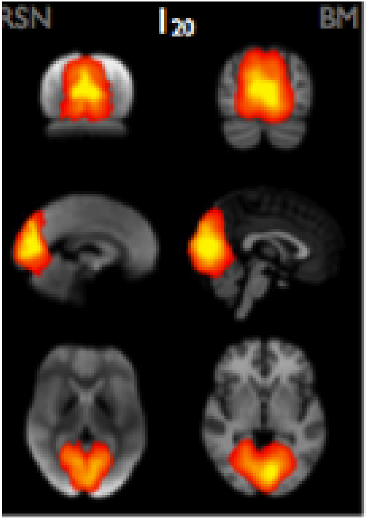
$$\begin{aligned} Y_t(i)|PaY_t(i) &= \mathbf{F}_t(PaY_t(i))\boldsymbol{\theta}_t(i) + v_t(i) \\ \boldsymbol{\theta}_t(i) &= G\boldsymbol{\theta}_t(i) + \mathbf{w}_t(i) \end{aligned}$$

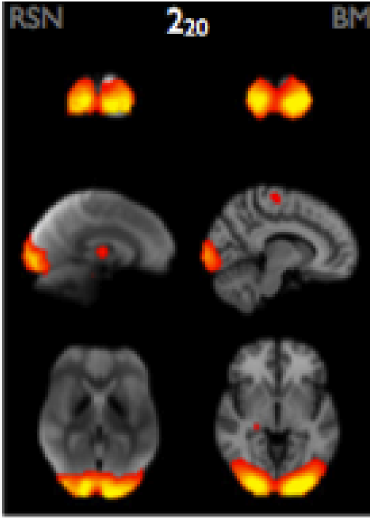
Example: fMRI regulation processes: structural elicitation

- 10 -20 regions of brain identified to give dynamic time series associated with each.
- Know that activation in one part of the brain can almost immediately excite (cause?) activity in another (measured by blood flow).
- It is well known that although directionality is enduring the strength of excitation drifts over time.
- Interest lies in directional network graph e.g. to check differences between healthy and diseased individuals.

Challenges: BN inappropriate ignores connectivity stochastics, but many competing plausible hypotheses.
Solution: Design a more bespoke graphical model of dependencies: here MDM works well as broad descriptor. Then embellish with other information.

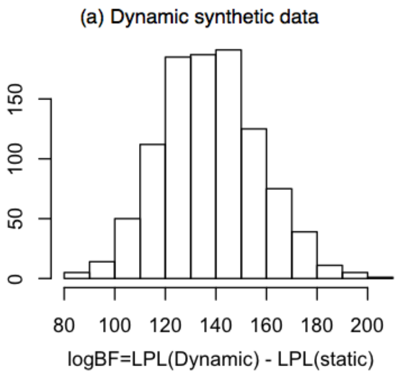
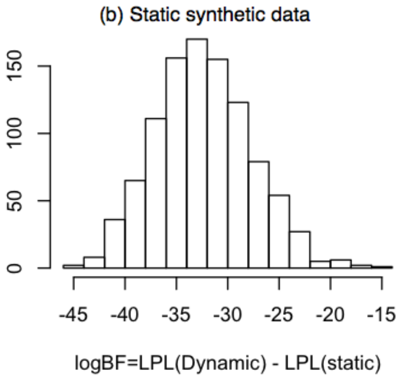
Two regions of the brain





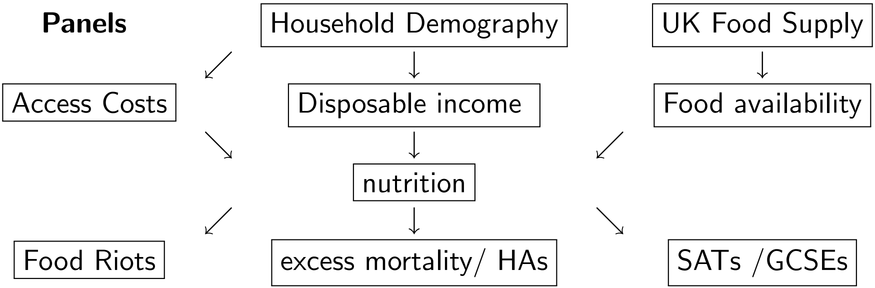
What we have done in this domain

- **Use LMDMs instead of BNs** as a graphical framework..
- Elicit **structural expert judgements** - symmetries, spatial,... to **preclude certain implausible flows**
- Supplement with quantitative priors - prior plausibility, strength of relationships & speed of change,
- **Discover** best supported models.
- Produce **diagnostics** & use expert judgments to improve fit (change points etc.)
- Feedback selection of mechanisms that seem well supported for expert sifting and further elicitation
- Compare our model with DCM (Costa et al 2012). Model search appears to work well not only on simulated data but also on various experiments.
- In simulations BF distinguish ME processes over the lengths of series we have. Selection from real data, appears scientifically plausible.


(a) Dynamic synthetic data
0
50
100
150
80
100
120
140
160
180
200
logBF=LPL(Dynamic) - LPL(static)

(b) Static synthetic data
0
50
100
150
-45
-40
-35
-30
-25
-20
-15
logBF=LPL(Dynamic) - LPL(static)

Combining judgments over different domains:Food Poverty

Huge no. of complex models describe different aspects of food system.



Panels agree overarching structure is MDM.

Question Can we paste this system together for decision support using MDM?

Answer Yes both formally & practically!!!

How we do this formally: Smith et al (2015) .

- A **rational** expected utility maximising *SupraBayesian (SB)* takes the agreed structural framework + conditional probability models from m panels of experts (often graphical). Nested graphs (summary, supergraph (MDM) vertex component subgraphs (MDMs, BNs, ..)
- An **agreed CK framework** helps graph define nature & composition of expert panels*:* associated with different components of the system. For each policy $d \in D$, based on own graphical model panel $G_i$delivers belief summaries $\{\Pi_i(d) : d \in D\}$, $i = 1.2, \ldots, m$ about parts of process within their expertise.
- Each domains has **varying complexity** & quality of information, all panels reason probabilistically: DBNs, MDMs.
- All participants **trust the expertise** of composing panels in a technical sense described in Barons et al (2017)
- SB then uses $\{\Pi_i(d) : d \in D\}$ to construct conditional expectations $\Pi = f(\Pi_1, \Pi_2, \ldots, \Pi_m)$ needed to **calculate her expected utilities** $\overline{U}(d)$ for each $d \in D$. These are then owned by everyone.

Practically: Leonelli & Smith(2015),Barons et al (2018)

Overarching MDM determines formulae for calculating expected utlities at any time for huge dynamic probabilistic structures having a wide variety of dependence relationships: see Leonelli & Smith (2015).

- Each panel need donate a **small no. of judgements** at each time step in terms of a predetermined st of conditional moments for each candidate policy. So dimension of a useful system can be surprisingly small!
- What these predetermined **moments** need to **depend on** topology of overarching **graph** & form of utilities.
- **Works for other customised probabilistic frames** as well (not just MDM)

Conclusion New nested Graphs supplemented by extra semantics of trust & acknowledgement of expertise expresses process accessibly

CEG's fit better than BN's (Barclay et al, 2013)

- Sometimes elicited information about how events might unfold.
- Then elicit event tree & colour edges the same if probabilities same. Transform into CEG (a coloured graph)
- Even when data are collected in complete tables, CEGs much better elicitation tool than BN!

$$\begin{tikzcd}
 & \text{Social Background} \arrow[dl] \arrow[d] \arrow[dr] & \\
\text{Economic Situation} \arrow[r] & \text{Family Life Events} \arrow[r] & \text{Hospital Admissions}
\end{tikzcd}$$

- With data can search. In study above then discover a CEG whose MAP score was 80 times better than this BN.

Search finds best explanation in terms of knock on effects.

$$
\begin{array}{ccccccc}
 & & & \text{low} & \text{HA}- & & \\
 & & & \nearrow & & \searrow\searrow & \\
 & \text{high} & \boxed{\text{LE}} & \rightrightarrows & \text{HA}^{=} & \rightrightarrows & w_\infty \\
 & \nearrow & & \nearrow\nearrow & \uparrow & & \upuparrows \\
\text{SB} & & \boxed{\text{LE}} & - & + & \rightarrow & \text{HA}^{+} \\
 & \searrow & \text{hi}\uparrow\text{gh} & & | & \nearrow\nearrow & \\
 & \text{low} & \text{ES} & \rightarrow & \text{LE} & & \\
 & & & \text{low} & & &
\end{array}
$$

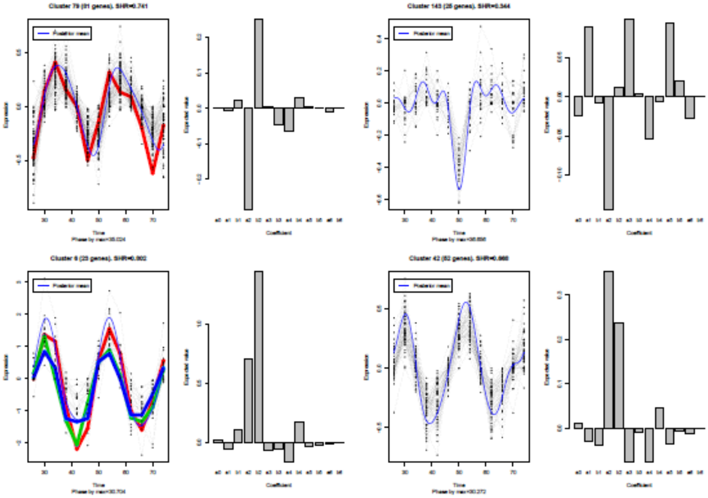
Can use the graph to feedback results for reflection: e.g.

- Economic S will not influence Life E's or child's hospital A's for families with high SB.
- High SB & low LE uniquely "causes" children a favourable HA^-.
- Low SB & low ES & then high or moderate life events lead to most child hospitalisation (access to credit?).

Example: regulatory graphs (Liverani & Smith, 2016)

Elicitation Interest circadian regulation. Regulation plausible if time course of regulated cluster relate in a particular way to time course of regulating cluster. e.g. short positive phase change (damping?)
Challenge Model space of scientifically plausible models massive $\backsim$ partition space of 20,000 different time courses
Solution Good data, with appropriate conjugate priors on Fourier basis gives product student t scores so greedy search feasible for independent cluster Usually end up with a partition of 100-200 interesting clusters of genes. Find families of posterior probabilities on each model and their estimated time courses to feed back to experts for further discrimination.

Gene Expression over time Liverani and Smith (2016)



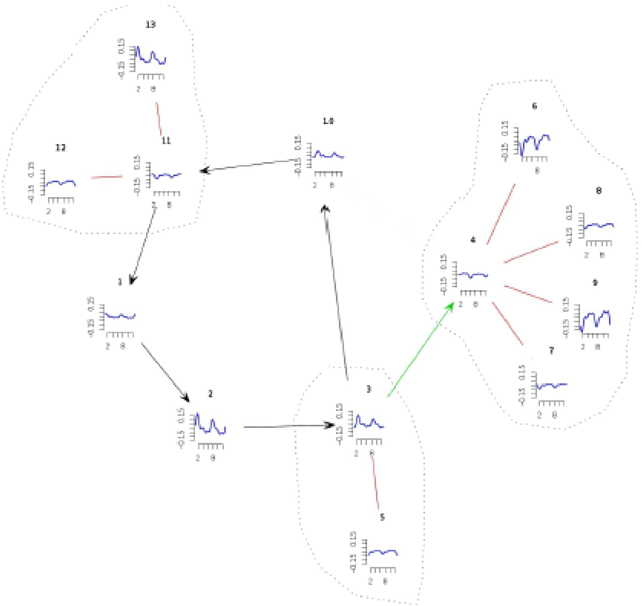
Problem

- Method assumes clusters independent but causal inferences we might want to make imply the parameters of two clusters are related.
- Standard DBN solutions lose conjugacy and allow too unstructured relationships over the time series (and relationships of the wrong form.
- Customise the graphs to represent dependences that are hypothesized exist and admit a causal interpretation.
- Our method does this and fast! It retains closed form scores.
- It also allows us to annotate the graph into a full and unambiguous probability model with an associated manipulative cause hypothesis

More specifically what we do

- Specify group of data transformations so transformed preprocessed clusters (within an orbit of the group) in same regulatory system: a set in same coarsen partition. Join two clusters into a *supracluster* if we can find parameters of a transformation which allows BF criteria to combine them.
- Clusters in same supraclusters hypothesised to be in same regulatory system.
- Cluster in same supracluster joined by directed connected graph where directions determined by parameters of transformation from one cluster into another.
- Existence of edge corresponds to a hypothesized scientifically plausible "causal" directionality within a connected component.
- Important features of process can be used to embellish the graphs.

MAP Regulation "Graph" for Supracluster of Aribidopsis



Conclusions

- Engaging in context or science encourages us to use various different forms of graphical representation to express expert judgements, draw out different logical consequences to examine & find best acts.
- Dangerous to use structures like BNs in unconsidered way. Worth developing bespoke graphs and supporting algebra.
- In particular causal hypotheses expressed in terms of faithful underlying belief structure.
- Great fun developing new formal graphical systems to help clients formulate & explore their particular brand of inferences.

Many thanks for your attention!!!!

Some references

Barons M, Wright S and Smith JQ, (2018)"Eliciting Probabilistic Judgments for Integrating Decision Support Systems", in: Dias LC, Morton A, Quigley J (eds), Elicitation: The science and art of structuring judgement. Springer, New York, Chapter 17 (to appear)

Collazo, R.A., Gorgen, C and Smith, J.Q.(2018) "Chain Event Graphs" Chapman Hall (to appear)

Leonelli, M. Roccomagno E., & Smith, J.Q. (2017) "A symbolic algebra for the computation of expected utilities in multiplicative influence diagrams" Annals of Mathematics and Artificial Intelligence (to appear)

Gorgen, C. & Smith, J.Q. (2017) "Equivalence classes of Staged Trees" Bernoulli (to appear)

Collazo, R.A. & Smith, J.Q.(2016) "A new family of Non-local Priors for Chain Event Graph model selection" Bayesian Analysis 11, no 4, 1165 - 12010-1309.

Some more references

Liverani, S. & J. Q. Smith. (2016) "Bayesian selection of graphical regulatory models". International J. of Approximate Reasoning, 77, 87–104. URL: doi:10.1016/j.ijar.2016.05.007

Smith J.Q., Barons, M.J. & Leonelli, M.(2015) "Bayesian Inference in Distributed Expert Systems" CRiSM report, ArXiv (under review)

Costa L, Smith,J.Q. Nicholls, T. and Cussens J. (2015) "Searching Multiregression Dynamic Models of Resting-State fMRI Networks Using Integer Programming" Bayesian Analysis Vol. 10, No. 2, 441-478

Leonelli. M., & Smith J.Q. (2015) "Bayesian Decision Support for Complex Systems with Many Distributed Experts " Annals of Operations Research DOI 10.1007/s10479-015-1957-7

Barclay, L.M. , Hutton, J.L. & Smith, J.Q.(2013) "Refining a Bayesian Network using a Chain Event Graph" International J. of Approximate Reasoning 54, 1300-1309.

- Useful trick: if priors on each candidate agree, then separation properties of multinomial or Gaussian data ensure

$$\log \pi_j - \log \pi_i = \sum_{I} \log \pi_i(\mathbf{x}) - \log \pi_i(\mathbf{x})$$

where I is the set of variables where they differ.

$$\begin{array}{cccccccccc} & & s & & M_i & & & s & & M_j \\ & \swarrow & & \searrow & & & \swarrow & \downarrow & \searrow & \\ e & \longrightarrow & f & \rightarrow & h & e & \longrightarrow & f & \rightarrow & h \end{array}$$

relative score

$$\log \pi_j(f|s,e) - \log \pi_i(f|e)$$

- So local search algorithms very quick (check space for 15 variables full search) &100's using greedy search

Selecting BN's (Barclay et al, 2012)

- Many methods available. My preference is Geiger and Heckerman(1997), who use local and global independnece and product Dirichlet priors giving log marginal likelihood score *linear* in cpts.

$$
\begin{tikzcd}
 & \text{Social Background} \arrow[dl] \arrow[d] \arrow[dr] & \\
\text{Economic Situation} \arrow[r] & \text{Family Life Events} \arrow[r] & \text{Hospital Admissions}
\end{tikzcd}
$$

- Search over all BNs is given above standard fits much better than original models. Note that edges as drawn fit causal development best.
- However several close competitors: where edges missing from ES$\longrightarrow$FLE, and one missing edge into HA.

Learning CEG's

- Prior stage floret independence is a generalisation of local and global independence in BNs. Just as in Geiger and Heckerman(1997), floret independence, together with appropriate Markov equivalence characterises this product Dirichlet prior (see Freeman and Smith, 2009)
- Choosing appropriate priors on model space and modular parameter priors over CEGs, for any CEG log marginal likelihood score is *linear* in stage components.
- Explicitly for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, let $s(\boldsymbol{\alpha}) = \log \Gamma(\sum_{i=1}^{k} \alpha_i)$ and $t(\boldsymbol{\alpha}) = \sum_{i=1}^{k} \log \Gamma(\alpha_i)$

$$\begin{aligned}
\Psi(C) &= \log p(C) = \sum_{u \in C} \Psi_{u(c)} \\
\Psi_{u(c)} &= \sum s(\alpha(i,u)) - s(\alpha^*(i,u)) + t^*(\alpha(i,u)) - t(\alpha(i,u))
\end{aligned}$$

Some other references

G. Shafer (2003). "The Art of Causal Conjecture". Cambridge, MA, MIT Press
Pearl, J., (2000). "Causality. Models, reasoning, and inference". Cambridge University Press, Cambridge
P. Spirtes, C. Glymour and R. Scheines (1993). "Causation, Prediction, and Search". Springer-Verlag, New York