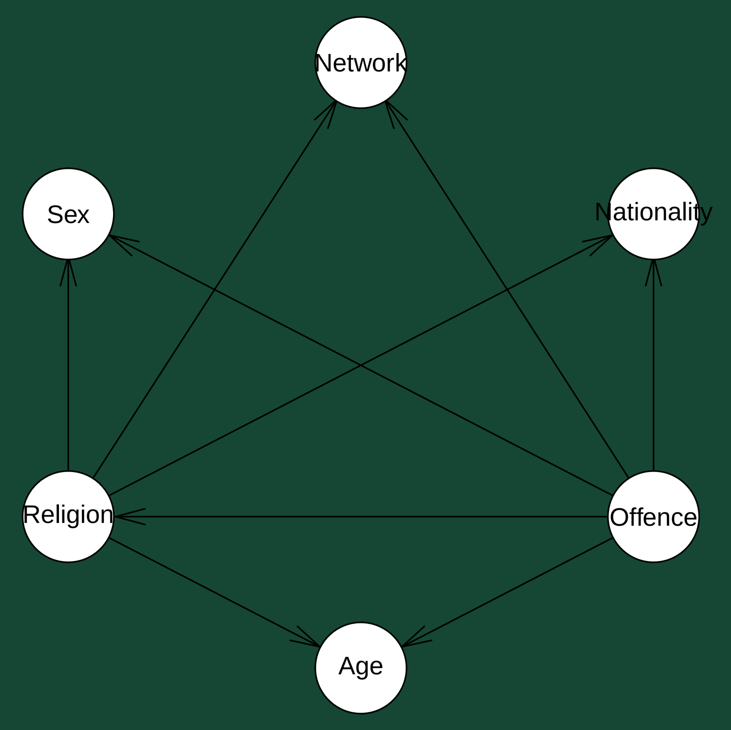
Data Description

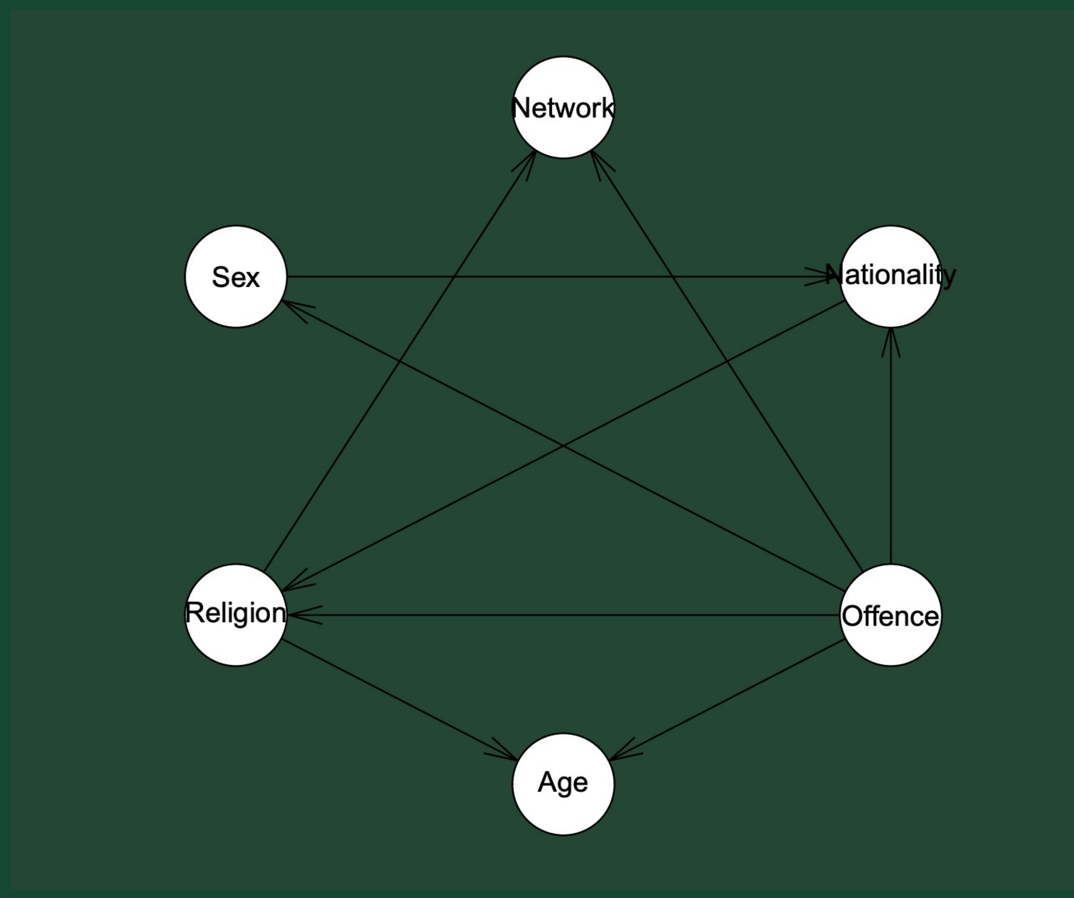
To illustrate the impact of using the beta divergence, we consider the example of radicalization in British prisons. The data was simulated from 1 in line with existing data from the Ministry of Justice on British prisons.
We used this example because it has many sparsely populated situations in the chain event graph. The CEG allows much more context-specific models.

The variables in the dataset are:

- **Sex**: Male (188), Female (12)
- **Religion**: religious (143), non-religious (51), not recorded (6)
- **Age**: young (84), old (65), medium (51)
- **Offence**: robbery (62), violence (61), others (29), drug (24), sexual offence (24)
- **Nationality**: British (178), foreigners (22)
- **Network**: rare (99), sometimes (74), always (27)



$\beta = 0.4$



$\beta = 0.5$

Beta Divergence for Model Selection in Bayesian Networks

Directionality across β

We search model structures for $\beta \in \{0.0, 0.1, \ldots, 1.0\}$ with `maxp=2`, starting from the empty network.
We focus on three variable pairs to illustrate how edge orientations shift as β varies:

- $R \sim N$: **Religion** vs. **Nationality**
- $S \sim N$: **Sex** vs. **Nationality**
- $S \sim R$: **Sex** vs. **Religion**

Legend: $A \to B$ | $B \to A$ | (no edge)

Pair	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$R \sim N$		$R \to N$	$R \to N$	$R \to N$	$R \to N$	$N \to R$	$N \to R$	$N \to R$	$N \to R$	$N \to R$	$N \to R$
$S \sim N$						$S \to N$	$S \to N$	$S \to N$	$S \to N$	$S \to N$	$S \to N$
$S \sim R$		$R \to S$	$R \to S$	$R \to S$	$R \to S$						

KL baseline vs. β-divergence
Structure changes across β

Abstract

We explore Beta divergence as a flexible scoring metric for Bayesian model selection, When $\beta = 0$, results recover the KL solution. As β increases, arc directionality can shift, revealing alternative structures. Shown here: the KL baseline, two β-settings where specific edges flip. the summery chart shows key arc changes across β.Well specified models are not sensitive to β but when we have poorly specified models in the M-ope n world, the β-divergence allows us to control the sensitivity to outliers.

β-divergence (prequential)

The β-divergence score $\ell^{(\beta)}(y_i; p) = \frac{1}{\beta} p(y_i)^{\beta} - \frac{1}{\beta+1} \Sigma y\, p(y)^{\beta+1}$ allows us to prequentially evaluate models

$$L(y; M_k) = \sum_{i=1}^{n} \ell^{(\beta)}\Big(y_i;\ p(\cdot \mid y_1, \ldots, y_{i-1}, M_k)\Big)$$

Credits

Roy Cannedy
Dr. Rachel Wilkerson (Baylor University)
Professor Jim Smith (University of Warwick)
Dr. Jack Jewson (Monash University)

References

Jewson, J., Smith, J.Q. and Holmes, C., 2018. *Principles of Bayesian inference using general divergence criteria*. **Entropy**, 20(6), p.442.

Collazo, R.A. and Smith, J.Q., 2016. *A new family of non-local priors for chain event graph model selection*. **Bayesian Analysis**, 11(4), pp.1165–1201. doi:10.1214/15-BA981.