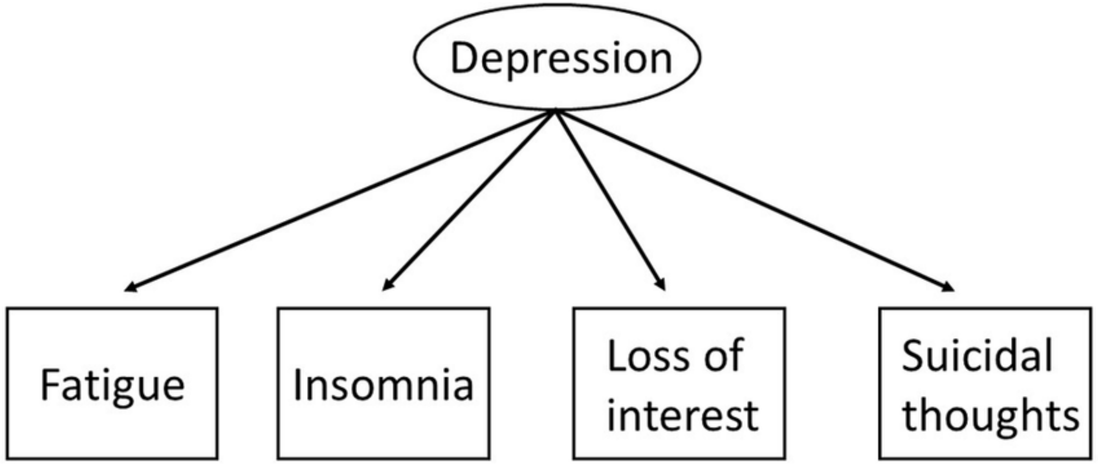
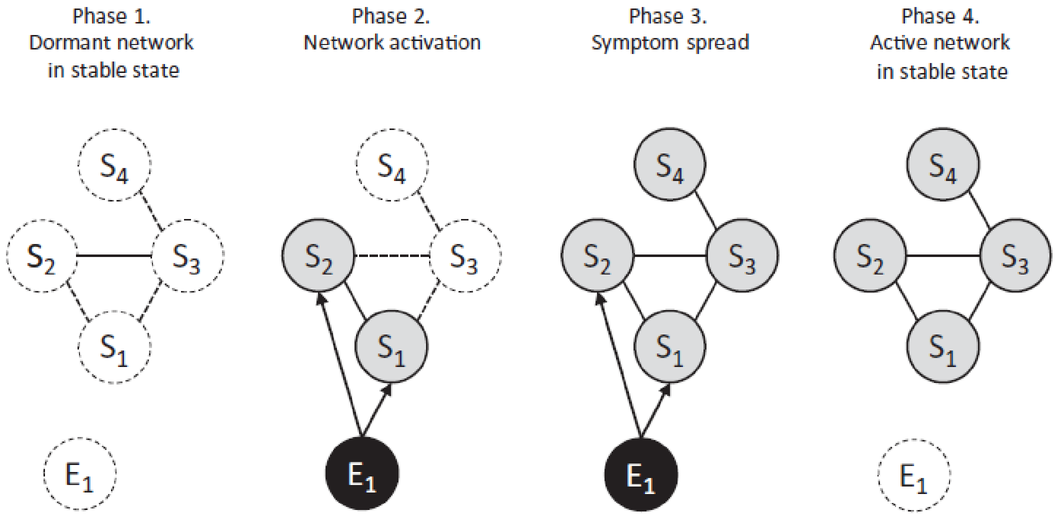
Unravelling the dynamics of bullying:
A Bayesian network and structural equation modelling approach

Andreas Bauer
Department of Preventive Medicine
Faculty of Medicine
University of São Paulo

Do symptoms share a hidden cause, or cause each other?





Latent variable models: an unobserved construct (e.g., depression) causes the observed symptoms; symptoms co-occur because they share the same latent cause.

Network models: symptoms influence each other directly; interactions can become self-sustaining.

Borsboom, 2017; doi:10.1002/wps.20375

A Tutorial on Bayesian Networks for Psychopathology Researchers

Giovanni Briganti[1, 2], Marco Scutari[3], and Richard J. McNally[1]
[1] Department of Psychology, Harvard University
[2] Laboratoire de Psychologie Médicale et Addictologie, Université libre de Bruxelles
[3] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland

Abstract

Bayesian Networks are probabilistic graphical models that represent conditional independence relationships among variables as a directed acyclic graph (DAG), where edges can be interpreted as causal effects connecting one causal symptom to an effect symptom. These models can help overcome one of the key limitations of partial correlation networks whose edges are undirected. This tutorial aims to introduce Bayesian Networks to identify admissible causal relationships in cross-sectional data, as well as how to estimate these models in R through three algorithm families with an empirical example data set of depressive symptoms. In addition, we discuss common problems and questions related to Bayesian networks. We recommend Bayesian networks be investigated to gain causal insight in psychological data.

Translational Abstract

In the last decade, the network framework for the study of mental disorders has emerged as a new way of investigating mental disorders as issuing from interactions among their constituent symptoms. Network analysis is the statistical aspect of this framework, as researchers use nodes (symptoms) and edges (connections between symptoms) to model disorders: Usually, network structures encode pairwise interactions among symptoms. In this study, we introduce Bayesian networks, models that can identify admissible causal relationships in cross-sectional data, as well as a tutorial for applied researchers on how to estimate those models in R. In addition, we discuss common problems and questions related to Bayesian network models.

Keywords: Bayesian networks, directed acyclic graphs, network modeling, causal inference, tutorial

Bayesian networks

A statistical model that represents a set of variables and their conditional dependencies using a Directed Acyclic Graph.

Directed Acyclic Graph (DAG)

- Directed = hypothesised causal effects
- Acyclic = no feedback loops

Briganti et al., 2022; doi: 10.1037/met0000479

Study characteristics

z-proso data from **K4** (11 years), **K5** (13 years), and **K6** (15 years)

Bullying perpetration and victimisation

- Social exclusion
- Verbal aggression
- Physical aggression
- Property destruction

Other influencing factors

- Competent conflict coping
- Internalising problems
- Delinquency (externalising problems)

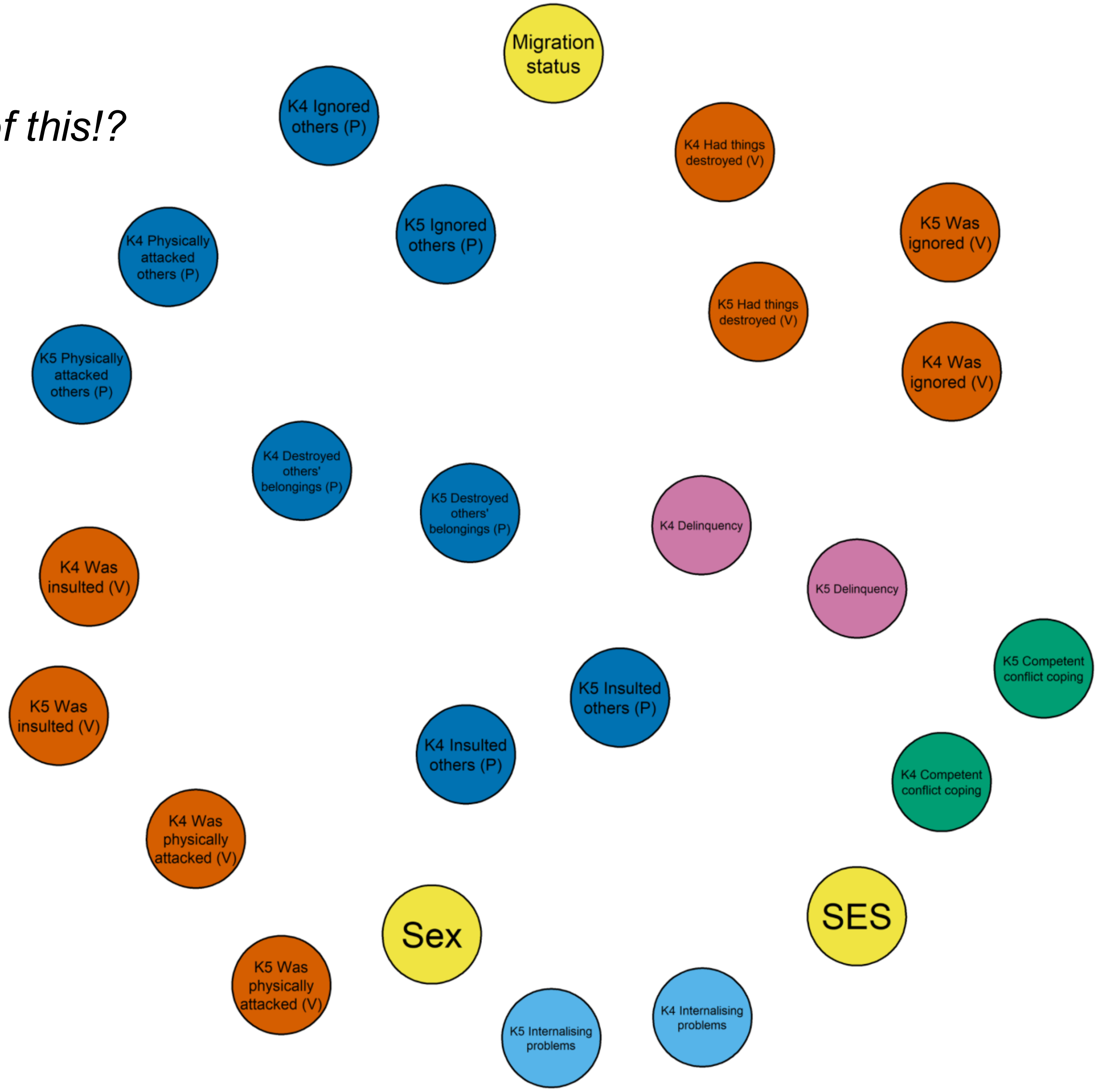
Time-invariant confounders

- Child sex
- Migration status
- SES

Research questions

- *How do specific bullying behaviours influence each other across time?*
- *What is the role of other influencing factors such as competent conflict coping, internalising problems, delinquency, and socio-demographic variables?*

How to make sense of this!?



Study workflow

1) Power analysis

- Via “Causal Discovery Power Calculator”
- A priori vs post-hoc
- Ideally: *training sample* (DAG – exploratory) and *testing sample* (SEM – confirmatory)

2) Missing data

- Unclear if more advanced missing data techniques are available (e.g., multiple imputation)
- Maybe limited to “single imputation”

3) Data preparation

- Split sample into two separate time windows (11-13 years and 13-15 years)
- No assumption of stationarity

4) DAG

- Use expert knowledge to define some paths
- Apply causal discovery algorithms to identify others
- Use bootstrapping to obtain the averaged network

5) SEM

- Estimate potential causal effects (based on your DAG)

6) Other metrics

- Identify communities + centrality metrics

Study workflow

1) Power analysis

- Via “Causal Discovery Power Calculator”
- A priori vs post-hoc
- Ideally: *training sample* (DAG – exploratory) and *testing sample* (SEM – confirmatory)

2) Missing data

- Unclear if more advanced missing data techniques are available (e.g., multiple imputation)
- Maybe limited to “single imputation”

3) Data preparation

- Split sample into two separate time windows (11-13 years and 13-15 years)
- No assumption of stationarity

4) DAG

- Use expert knowledge to define some paths
- Apply causal discovery algorithms to identify others
- Use bootstrapping to obtain the averaged network

5) SEM

- Estimate potential causal effects (based on your DAG)

6) Other metrics

- Identify communities + centrality metrics

Preliminary analyses

Expert knowledge

Blacklist
(never included)

Whitelist
(always included)

No paths to time-invariant confounders
(e.g., bullying → sex)

+

No paths between time-invariant confounders
(e.g., SES → sex)

+

No paths from wave 5 to wave 4 (future → past)

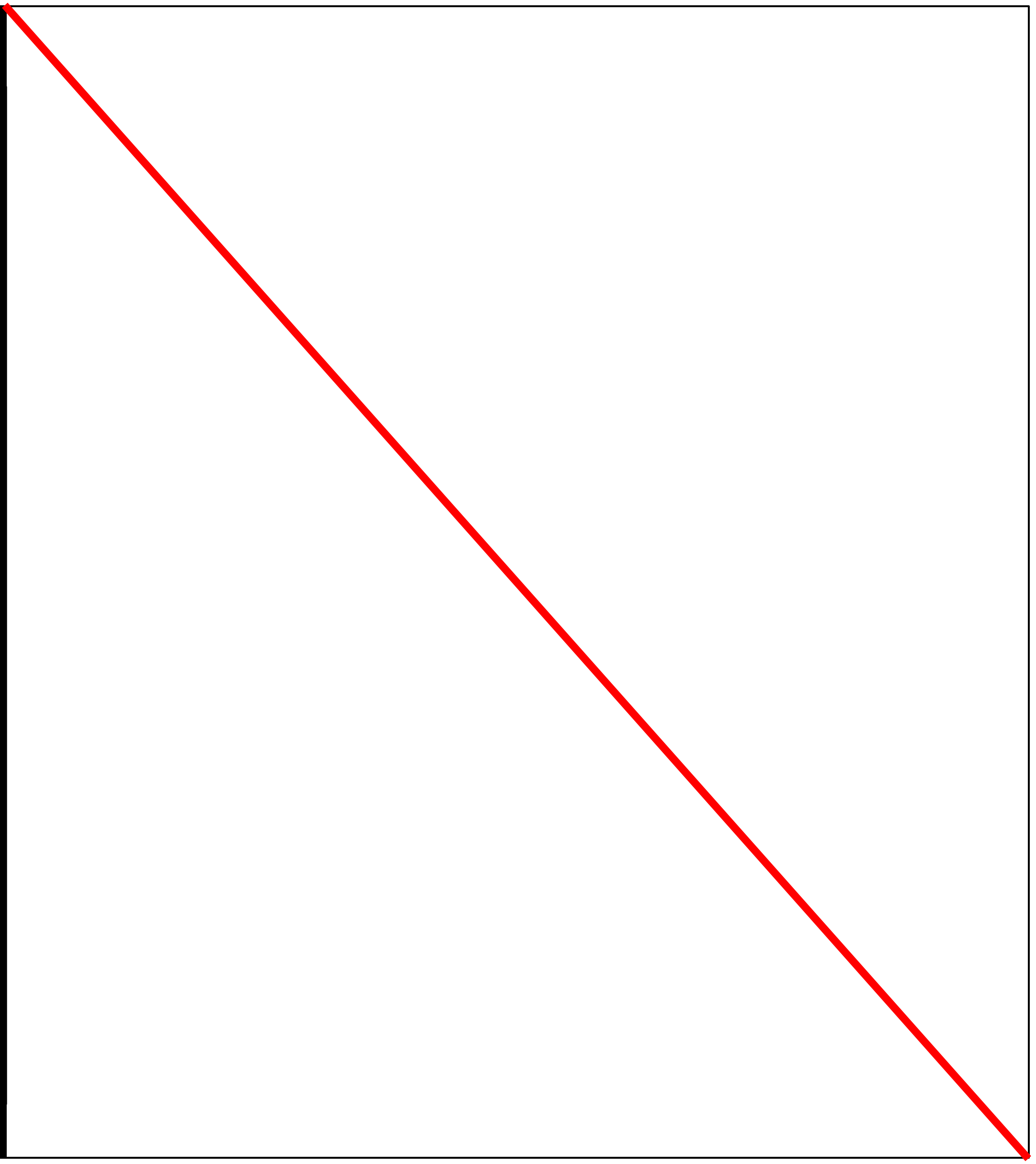
+

No contemporaneous edges within wave 4
(avoids feedback loops)

+

No contemporaneous edges within wave 5
(avoids feedback loops)

= 413 paths to blacklist in our model



The three main algorithm families in causal discovery

Constraint-based learning algorithms

- Rely on *conditional independence tests* to decide whether edges should be present (i.e., are two variables still dependent after conditioning on a set of others?)
- For example: PC (pc.stable), Grow-Shrink (gs)

Score-based learning algorithms

- Assign a numerical *score* to each candidate network (e.g., BIC) – keeps the one with the best score
- For example: Hill-Climbing (hc), Tabu Search (tabu)

Hybrid learning algorithms

- Use conditional independence tests to reduce the number of candidate networks, then select the highest-scoring network
- For example: Max-Min Hill-Climbing (mmhc), Restricted Maximisation (rsmax2)

Briganti et al., 2022; doi: 10.1037/met0000479

The three main algorithm families in causal discovery

Constraint-based learning algorithms

- Rely on *conditional independence tests* to decide whether edges should be present (i.e., are two variables still dependent after conditioning on a set of others?)
- For example: PC (pc.stable), Grow-Shrink (gs)

Score-based learning algorithms

- Assign a numerical *score* to each candidate network (e.g., BIC) – keeps the one with the best score
- For example: Hill-Climbing (hc), Tabu Search (tabu)

Hybrid learning algorithms

- Use conditional independence tests to reduce the number of candidate networks, then select the highest-scoring network
- For example: Max-Min Hill-Climbing (mmhc), Restricted Maximisation (rsmax2)

No single algorithm family (or individual algorithm) consistently outperforms the others!

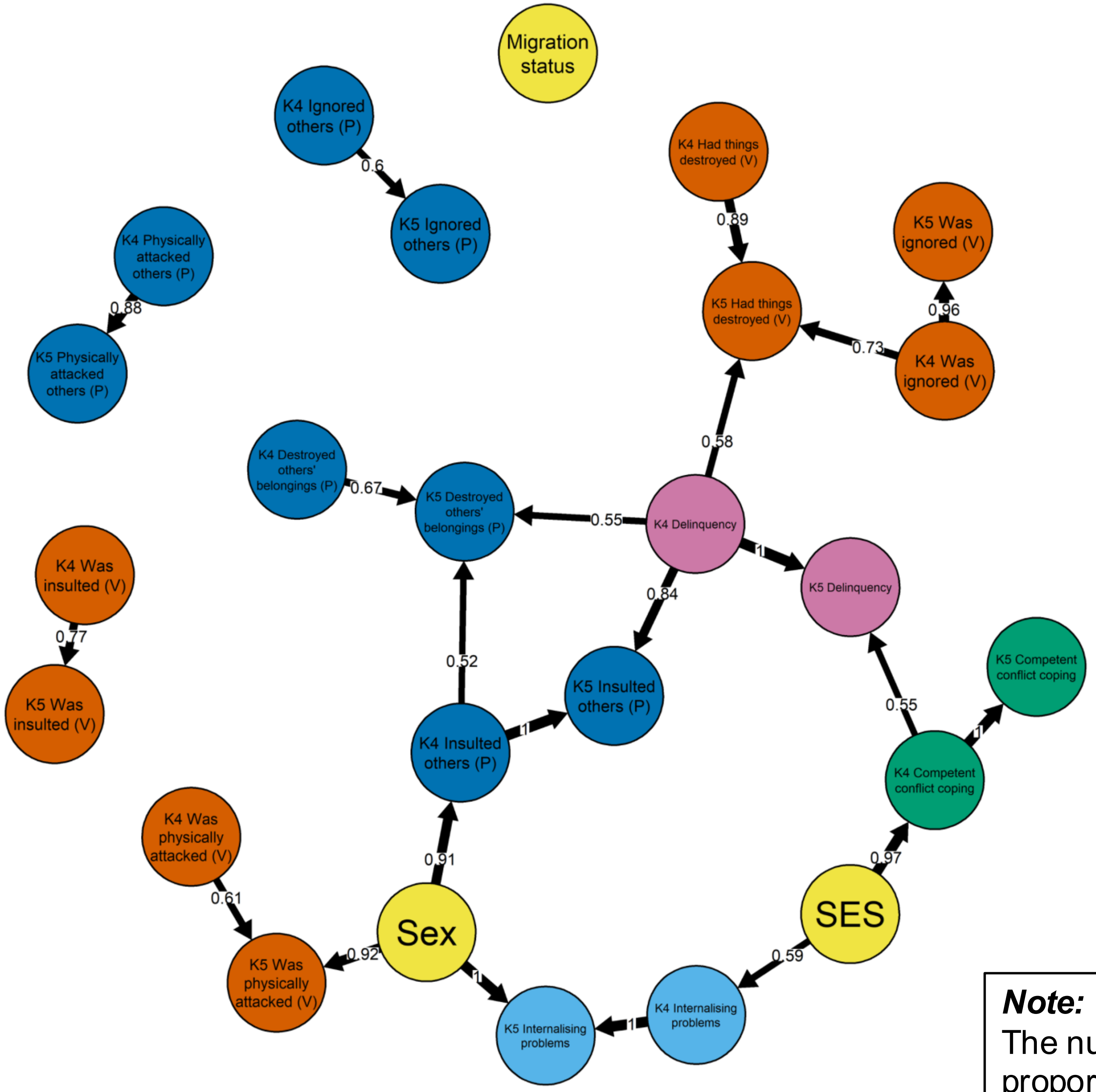
Best practice is to apply at least one algorithm from each family for comparison…

Briganti et al., 2022; doi: 10.1037/met0000479

Let's use our ***blacklist*** and the ***pc.stable algorithm*** to estimate the DAG.

Let's do that a 1,000 times (***bootstrapping***) …

We apply an empirically-derived threshold to get the ***averaged network***: edges that were replicated in at least 48% of bootstrap samples are retained



Note:
The numbers on arrows show the proportion of bootstrap samples in which the edge appeared.

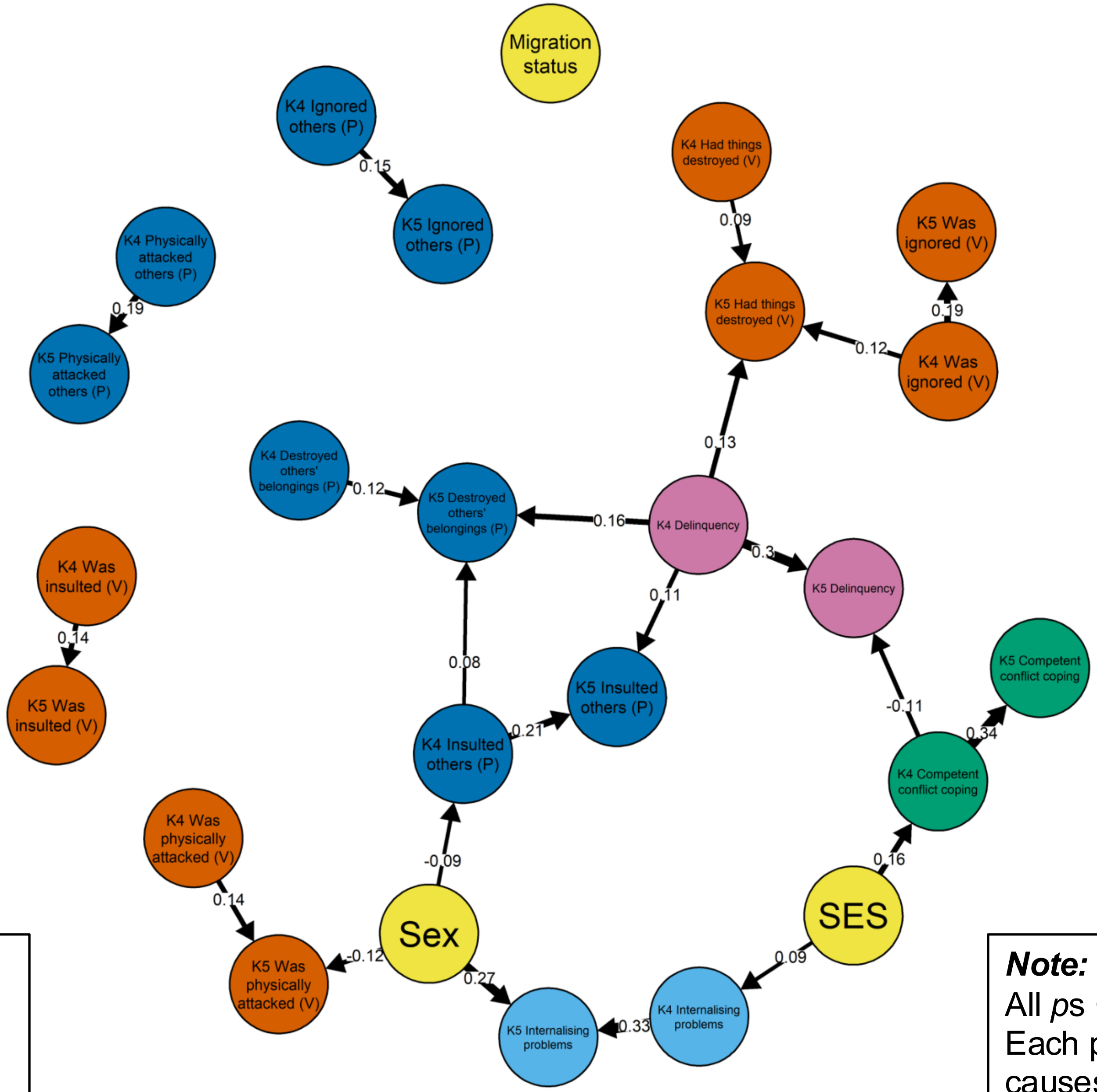
Scutari & Nagarajan, 2013; doi:10.1016/j.artmed.2012.12.006

from	to	strength	direction
K4_ConfCopComp	K5_ConfCopComp	1.000	0.9780000
K4_D_mean	K5_D_mean	1.000	0.9990000
K4_nSBQ_ANXDEP	K5_nSBQ_ANXDEP	1.000	0.9930000
sex	K5_nSBQ_ANXDEP	1.000	0.9935000
K4_insP	K5_insP	0.998	0.9834669
SES	K4_ConfCopComp	0.968	1.0000000
K4_ignV	K5_ignV	0.963	1.0000000
sex	K5_attV	0.919	0.9989119
sex	K4_insP	0.906	1.0000000
K4_destV	K5_destV	0.888	1.0000000
K4_attP	K5_attP	0.881	0.9818388
K4_D_mean	K5_insP	0.839	1.0000000
K4_insV	K5_insV	0.766	1.0000000
K4_ignV	K5_destV	0.733	1.0000000
K4_destP	K5_destP	0.673	0.9992571
K4_attV	K5_attV	0.611	1.0000000
K4_ignP	K5_ignP	0.601	0.9958403
SES	K4_nSBQ_ANXDEP	0.592	1.0000000
K4_D_mean	K5_destV	0.583	1.0000000
K4_D_mean	K5_destP	0.552	1.0000000
K4_ConfCopComp	K5_D_mean	0.547	1.0000000
K4_insP	K5_destP	0.525	0.9990476

For some edges, there is *uncertainty* whether the edge exists (strength < 1).

When edges are present, their direction is estimated with *high certainty* (direction ≈ 1).

Standardised regression coefficients in the DAG-based SEM



Good model fit
CFI = 0.97
TLI = 0.94
RMSEA (90% CI) = 0.04 (0.03-0.05)
SRMR = 0.07

Note:
All *ps* < 0.05.
Each path was adjusted for its direct causes; residuals of variables at the same time point are allowed to covary.

Preliminary conclusions

- **Stability** from ages 11 to 13 years across all domains (autoregressive effects)
- Weak evidence for **cross-bullying influences**: verbal aggression (P) → property destruction (P); social exclusion (V) → property destruction (V)
- **Delinquency** may play a more central role in increasing bullying perpetration and victimisation – *what about conceptual overlap?*
- **Internalising problems** and **competent conflict coping** did not show direct longitudinal effects on bullying, once stability, confounders and contemporaneous associations were accounted for
- **Sex** and **SES**, but not migration status, showed some effects on bullying and other influencing factors

Take-home message

We can use Bayesian network modelling to identify a DAG, which is then tested and quantified via SEM.

Thank you!