Exploratory network analysis of large social science questionnaires

Robert J. B. Goudie
Department of Statistics
University of Warwick
Coventry, UK

Sach Mukherjee
Department of Statistics &
Centre for Complexity Science
University of Warwick
Coventry, UK

Frances Griffiths
Health Sciences Research Institute
University of Warwick
Coventry, UK

Abstract

There are now many large surveys of individuals that include questions covering a wide range of behaviours. We investigate longitudinal data from the Add Health survey of adolescents in the US. We describe how structural inference for (dynamic) Bayesian networks can be used to explore relationships between variables in such data and present this information in an interpretable format for subject-matter practitioners. Surveys such as this often have a large sample-size, which, whilst increasing the precision of inference, may mean that the posterior distribution over Bayesian networks (or graphs) is concentrated on disparate graphs. In such situations, the standard MC^3 sampler converges very slowly to the posterior distribution. Instead, we use a Gibbs sampler (1), which moves more freely through graph space. We present and discuss the resulting Bayesian network, focusing on depression, and provide estimates of how different variables affect the probability of depression via the overall probabilistic structure given by the Bayesian network.

1 INTRODUCTION

Hypotheses of multifactorial causes of symptoms and outcomes play an important role in the social sciences and in public health. Regression-based approaches are widely-used in these fields to explore such hypotheses. A great deal of insight can be gained through such approaches, but it is sometimes overly constraining to fix a particular quantity as the dependent variable, especially if the goal is to explore the possibility of unexpected relationships between the data. Instead, we can consider a number of variables on an equal footing, and study the possibility of unexpected relationships in the data.

Graphical models provide a statistical framework within which the relationship between variables can be studied. These models enable complex multivariate distributions to be decomposed into simpler local distributions. This can reveal a great deal about the relationships between the variables, as well as provide a statistical and computationally tractable description of their (often large) joint distribution. The decomposition is formed by the conditional independence structure, which can be represented by a graph. The use of graphs helps to make the interpretation of the model simpler. In this paper, we focus on the structure of the model, as given by the graph. We aim to make inference about this using statistical model selection. The structure of the model suggests how the different components of the system interact, which may be helpful in understanding the system as a whole. These methods have been widely adopted in molecular biology (2, 3), and have been used in some areas of medical sciences (4).

Consideration of unexpected relationships between factors requires datasets that incorporate a wide range of topics. Such data is now widely available for representative samples of populations in many countries, and for many sub-groups of interest. Many of these datasets are derived from surveys that are general in scope, and are not collected to study any one particular question. For example, in the US, the health of the whole population is representatively sampled annually for the Behavioral Risk Factor Surveillance System (BRFSS) survey, and the Add Health study, which we use here, followed a cohort of young people from 1994 until 2008. Data from both of these have been used in scores of studies, but these commonly focus on one specific aspect, often using the data to evaluate existing hypotheses. Given the wide scope inherent in the design of these studies and the large samples available in many cases, it is possible to broaden the scope of the analysis by considering richer

structures. In this paper, we discuss the potential that such a more explorative approach yields. We do not seek to make conclusive causal claims, but instead suggest that a broader approach may uncover important aspects that have been neglected.

Our focus will be on depression among adolescents in the US, drawing on data from the National Longitudinal Study of Adolescent Health (Add Health). It is estimated that around 1–6% of adolescents each year are affected by depression (5, 6). The effects of depression in this age-group are wide-ranging (7), and include the stigma associated with poor mental health more generally (8). There is considerable evidence that there are a wide range of causal factors for depression amongst adolescents, spanning biological, psychological and social domains. Understanding these causal factors and separating them from the consequences of depression has been recognised as an important aim (9). Some of the relevant causal factors may interact and the approach taken here accounts for this.

The remainder of this paper is organised as follows. We first introduce the AddHealth dataset and describe the Bayesian network framework. Inference for Bayesian networks is performed using Markov Chain Monte Carlo (MCMC), but the large sample size of the dataset we consider makes achieving convergence difficult because the posterior distribution may be concentrated on disparate graphs, and so we describe an alternative sampler that has superior properties in this situation. Whilst the PC-algorithm (10, 11) has properties that often make it attractive in such contexts, we found that the results in this situation were not robust (see Discussion). We then present and discuss the results for the Add Health dataset.

2 MATERIALS AND METHODS

2.1 Add Health

The data that we use are drawn from the National Longitudinal Study of Adolescent Health (Add Health) that explores health-related behavior of adolescents (12) in the US. The questionnaire contains over 2000 questions that cover many aspects of adolescent behaviours and attitudes. We consider the representative sample of adolescents from Waves I and II of the in-home section, and the parental questionnaire from Wave I of the study. The analysis we perform is not feasible when the data is not complete (see Discussion), and so individuals with missing data were removed from the study. Removing incomplete samples leaves 5975 individuals in the study.

Our measure of depression is a self-assessed scale based upon the Centre for Epidemiologic Studies Depression Scale (CES-D) (13). Two questions from the 20-item scale are omitted from AddHealth, and two are modified, and so we scale the score given by the available questions (14). A Receiver Operating Characteristic (ROC) analysis showed that thresholds of 24 for females and 22 for males provided the best agreement with clinical assessments of depression (15). We use this threshold to create a binary indicator of depression status.

Many of the remainder of the variables that we consider (Table 1) are drawn from the risk factors described in the depression literature, and the mental health literature more generally. A recent review (8) described a wide range of factors that are associated with poor mental health in young people, including gender, poverty, violence and the absence of social networks in the local neighbourhood. The quality of relationships with parents is also thought to be important, especially with the mother (16), as are parental alcohol problems (17) and parental discord (16). The individual's use of alcohol, drugs, smoking and HIV/AIDS are all also associated with depression (18, 19). Physical exercise has been proposed in some studies as a useful intervention for the management of depression, but many of these studies have been deemed to be poor quality (20).

2.2 Bayesian Networks

Our study uses Bayesian networks to explore the relationships between variables in the Add Health study. Bayesian networks are a particular type of graphical model that enable classes of probability distributions to be specified using a directed acyclic graph (DAG). A Bayesian network G is represented using a DAG with vertices $V = (V_1, \ldots, V_p)$, and directed edges $E \subset V \times V$. The vertices correspond to the components of a random vector $\mathbf{X} = [X_1, \ldots, X_p]^T$, subsets of which will be denoted by X_A for sets $A \subseteq \{1, \ldots, p\}$. For $1 \leq i, j \leq p$, we define the parents G_j of each node V_j to be the subset of vertices V such that $V_i \in G_j \Leftrightarrow (V_i, V_j) \in E$. Specifying the parents of the vertices determines the edges E of the graph G. We denote by $\mathcal{G}$ the space of all possible directed acyclic graphs with p vertices. We will use X_{G_i} to refer to the random variables that are parents of X_i in the graph G.

The graph specifies that the joint distribution for $\mathbf{X}$, with parameters $\theta = (\theta_1, \ldots, \theta_p)$, can be written as a product of conditional distributions $p(X_i \mid X_{G_i}, \theta_i)$, given the variables X_{G_i} corresponding to the parents of X_i in the graph.

$$p(\mathbf{X} \mid G, \theta) = \prod_{i=1}^{p} p(X_i \mid X_{G_i}, \theta_i)$$

We will need to be able to evaluate the marginal likelihood $p(\mathbf{X} \mid G)$ easily, and so we consider only a conjugate analysis in which the conditional distributions $p(X_i \mid X_{G_i}, \theta_i)$ are multinomial, with Dirichlet priors $p(\theta_i)$ for each θ_i. In this case, the marginal likelihood can be evaluated analytically. Suppose each X_i takes one of r_i values, and define q_i as the number of levels of the sample space of X_{G_i}, each element of which we call a configuration. For each configuration j of X_{G_i}, let N_{ijk} be the number of observations in which X_i takes value k. We assume the Dirichlet priors for each θ_i, each with hyperparameters N'_{ijk}, are independent. We define $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ and $N'_{ij} = \sum_{k=1}^{r_i} N'_{ijk}$, and the local score $p(X_i \mid X_{G_i})$ to be

$$p(X_i \mid X_{G_i}) = \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})}.$$

The marginal likelihood can be shown to equal the product $p(\mathbf{X} \mid G) = \prod_{i=1}^{p} p(X_i \mid X_{G_i})$ of these local scores (21).

2.3 Structural inference for Bayesian Networks

We aim to make inference about the DAG G, given data $\mathbf{X}$ and so our interest focuses on the posterior distribution $\Pr(G \mid \mathbf{X})$ on Bayesian networks. Under the assumptions we have made, this can be written in terms of the marginal likelihood $p(\mathbf{X} \mid G)$, and a prior $\pi(G)$ for the Bayesian network structure.

$$\Pr(G \mid \mathbf{X}) \propto \pi(G) \prod_{i=1}^{p} p(X_i \mid X_{G_i})$$

The priors $\pi(G)$ can be chosen to encode domain information (3). For the analyses in this paper, we choose an improper prior $\pi(G) \propto 1$ that is flat across the space of graphs.

The posterior distribution $\Pr(G \mid \mathbf{X})$ is difficult to evaluate, because cardinality of $\mathcal{G}$ grows super-exponentially in p. This motivates the use of approximations to $\Pr(G \mid \mathbf{X})$, which are usually based on Markov chain Monte Carlo (MCMC).

2.4 Approximate inference for Bayesian Networks

The standard form of MCMC that is used for structural inference for Bayesian networks is MC^3 (22). This is a Metropolis-Hastings sampler that explores $\mathcal{G}$ by proposing to add or remove a single edge from the current graph G. This sampler works surprisingly well in many situations, but if the posterior distribution is not unimodal, the local moves may fail to explore the space fully because the sampler may become 'trapped' in one mode. This issue becomes more severe as the sample size increases because the posterior distribution becomes more concentrated. A natural approach in such situations is to use the PC-algorithm (10, 11), which has been shown to be asymptotically consistent (23), but we found in this case that the results were not robust (see Discussion).

Our analyses in this paper were performed using a Gibbs sampler (1), which we found to converge rapidly to its equilibrium state. A naïve Gibbs sampler for structural inference that proposes single-edge additions and removals can easily be constructed, but this sampler offers no advantages over the analogous MC^3. This naïve scheme, however, can be improved by 'blocking' together a number of components, and sampling from their joint conditional distribution. In theory, any group of components can be taken as a block, but sampling from their joint conditional distribution needs to be possible and, ideally, computationally quick.

For Bayesian networks, the most natural blocks are those consisting of parent sets $G_1, \ldots, G_p$. This is natural because the marginal likelihood $p(\mathbf{X} \mid G)$ for a graph G factorises across vertices into conditionals $p(X_j \mid X_{G_j})$ and these conditionals depend on the parent set of the vertex. Therefore, since any graph $G \in \mathcal{G}$ can be specified by a vector $G = (G_1, \ldots, G_p)$ of parent sets, the posterior distribution on Bayesian networks $G \in \mathcal{G}$ can be written as functions of $G_1, \ldots G_p$ in the following way.

$$\Pr(G_1, \ldots, G_p \mid \mathbf{X}) \propto \pi(G_1, \ldots, G_p) \prod_{i=1}^{p} p(X_i \mid X_{G_i})$$

In the following, we will denote subsets of the vector $G = (G_1, \ldots, G_p)$ by $G_A = \{G_k : k \in A\}$, and the subset given by the complement $A^C = \{1, \ldots, p\} \setminus A$ of a set A will be denoted by $G_{-A} = \{G_k : k \in A^C\}$. In particular, the complete graph can be specified by $G = (G_1, \ldots G_p) = (G_i, G_{-i})$ for any $i \in \{1, \ldots, p\}$.

To be able to construct a Gibbs sampler using parent sets, we need to find their conditional distribution, given the other parent sets $G_{-j} = \{G_1, \ldots, G_{j-1}, G_{j+1}, \ldots, G_p\}$. Parent sets G_j for which $G = (G_j, G_{-j})$ is cyclic will have no probability mass in the conditional distribution. Let $K_j^\star$ be the set of parent sets G_j such that $G = (G_j, G_{-j})$ is acyclic. The conditional posterior distribution of G_j is multinomial, with weights given by the posterior distribution of $G = (G_j, G_{-j})$. When the cardinality of $K_j^\star$ is constrained (for example, by restricting the maximum number of parents of each node) the conditional posterior distribution for $G_j \in K_j^\star$ can be evaluated

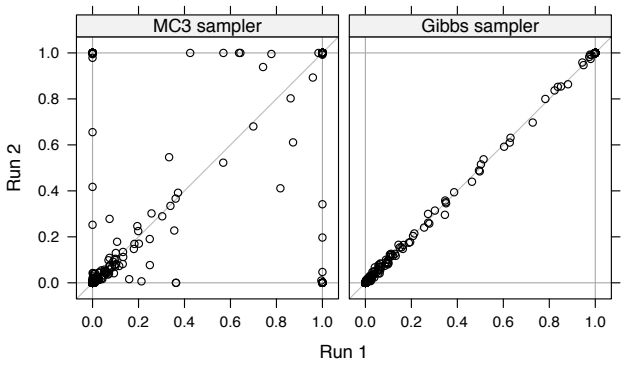

Figure 1: Diagnostic runs for MC^3 (left) and the Gibbs sampler (right). The posterior edge probabilities given by two independent runs are plotted against each other. When the two runs give the same estimates of the posterior edge probabilities, all of the points appear on the line $y = x$. We observe that the two Gibbs runs gives similar posterior edge probabilities, but the MC^3 runs do not. (5 runs of 750,000 samples (MC^3) or 100,000 samples (Gibbs) of each sampler were performed; the first half of the samples were discarded as burn-in; mean Pearson correlation between runs was 0.9999 ± 0.0002 (standard deviation) for Gibbs and 0.6322 ± 0.0477 for MC^3.)

exactly.

$$\Pr(G_j \mid G_{-j}, \mathbf{X}) = \frac{\Pr(G_j, G_{-j} \mid \mathbf{X})}{\Pr(G_{-j} \mid \mathbf{X})} = \frac{\Pr(G_j, G_{-j} \mid \mathbf{X})}{\sum_{G_j \in K_j^\star} \Pr(G_j, G_{-j} \mid \mathbf{X})} \quad (1)$$

We can improve the speed of convergence of this sampler by allowing pairs of parent sets to be sampled together. At each step of the Gibbs sampler we conditionally sample pairs of parent sets (G_{j_1}, G_{j_2}), given the remainder of the graph $G_{-\{j_1,j_2\}}$. Parent sets $G_{-\{j_1,j_2\}}$ such that $G = (G_{j_1}, G_{j_2}, G_{-\{j_1,j_2\}})$ is cyclic have no probability mass in the conditional distribution. Let $K^\star_{j_1,j_2}$ be the set of pairs of parent sets (G_{j_1}, G_{j_2}) such that $G = (G_{j_1}, G_{j_2}, G_{-\{j_1,j_2\}})$ is acyclic. For $(G_{j_1}, G_{j_2}) \in K^\star_{j_1,j_2}$, the conditional posterior distribution is multinomial, by analogy with (1), with weights given by posterior distribution of $G = (G_{j_1}, G_{j_2}, G_{-\{j_1,j_2\}})$.

$$\Pr(G_{j_1}, G_{j_2} \mid G_{-\{j_1,j_2\}}, \mathbf{X}) = \frac{\Pr(G_{j_1}, G_{j_2}, G_{-\{j_1,j_2\}} \mid \mathbf{X})}{\sum_{(G_{j_1}, G_{j_2}) \in K^\star_{j_1,j_2}} \Pr(G_{j_1}, G_{j_2}, G_{-\{j_1,j_2\}} \mid \mathbf{X})}$$

Similarly, sets of three parent sets can be conditionally sampled. Full technical details are presented in (1).

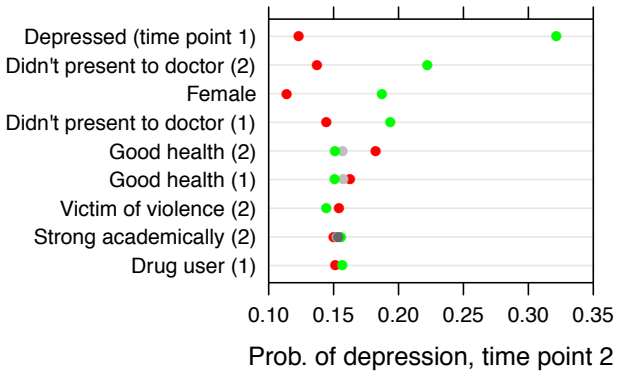

Figure 3: Conditional probability of depression. The conditional probability of being depressed at Wave II given the variable indicated is changed to the level indicated by the colours, conditional on the DAG shown in Figure 2. For binary variables, ● is true, and ● is false; shades of grey indicate intermediate levels. Wave number (time point) is indicated in parentheses. Only variables for which the conditional probability differed between levels by at least 0.005 are displayed.

3 RESULTS

The variables that we consider are detailed in Table 1. As is common when using graphical models (24), all of these variables were grouped, initially into 'Background', 'Wave I' and 'Wave II', and then refined into whether the question asked about the long- or short-term, as shown in Table 2. These groups define constraints on the Bayesian networks that are considered. Specifically, no edges can be directed backwards through the groups. Edges, however, are allowed within groups. For example, no edge is allowed to be directed into 'Gender', and no edge can pass backwards in time, for example, from Depression at Wave II to Depression at Wave I. Additionally, no edge can pass from a short-term variable to a long-term variable, for example, from Depressed at Wave I to Have HIV/AIDS at Wave I.

We precomputed the local scores, and then drew 100,000 samples (the first half of which were discarded as burn-in) using the Gibbs sampler (Section 2.3), which took 30 minutes (on a single core of a cluster computer). The graph space was constrained such that no node had more than 3 parents, to ensure Equation 1 could be evaluated.

We ran 5 independent samplers, with disparate initial states. This enables a simple test of convergence to be performed that compares the posterior edge probabilities obtained from each of the independent runs (25). The agreement between runs can be examined graphically by plotting the edge probabilities against

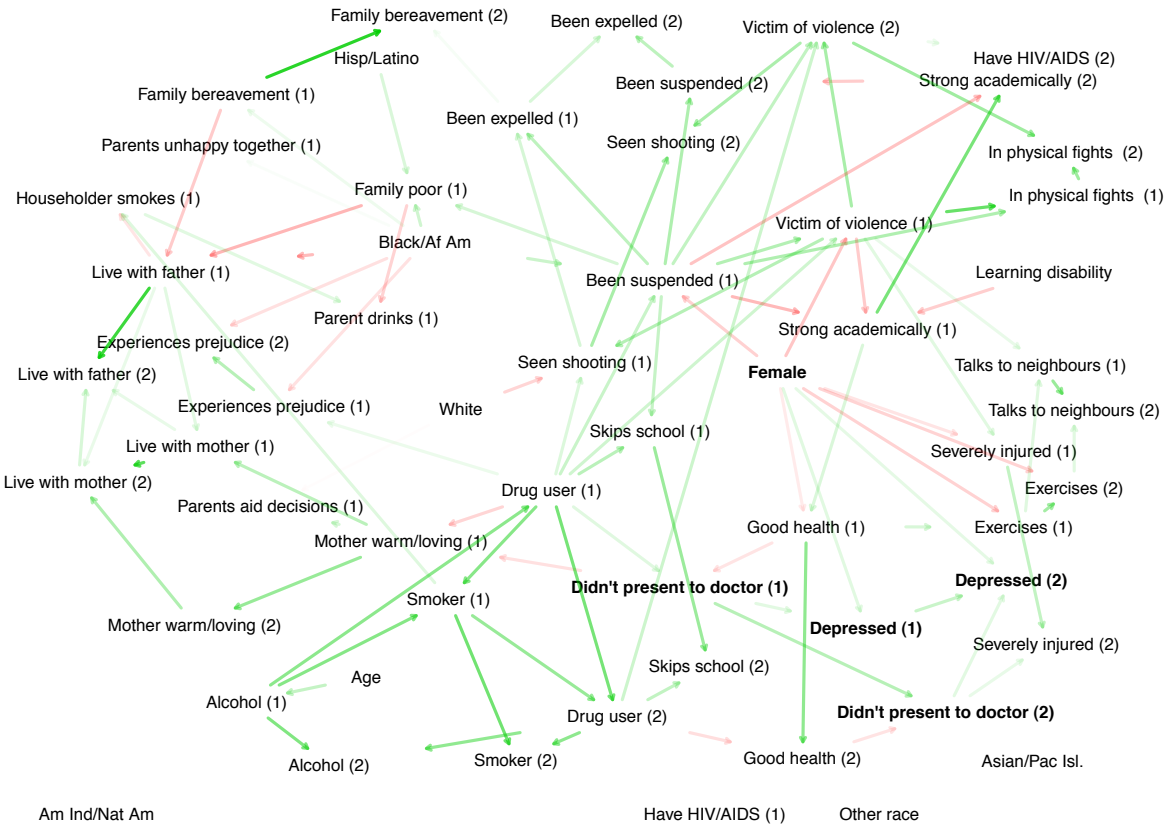

Figure 2: Summary network for the AddHealth variables considered. The edge colors are given by the Kendall correlation coefficents between the two variables, with green edges corresponding to positive correlation, and red edges to negative correlation. The strength of the correlation is indicated by the transparency of the line, with greater transparency indicating weaker correlation. The variables 'Depressed (1)', 'Depressed (2)' and their parents are shown in bold.

each other (Figure 1). Mean Pearson correlation coefficients between edge probabilities from pairs of runs were 0.9999 ± 0.0002 (standard deviation) for the Gibbs sampler and 0.6322 ± 0.0477 for MC^3. The agreement between the independent runs of the Gibbs sampler gave us confidence in our results, in contrast to the large disagreements between MC^3 runs. In addition, cumulative edge probability plots for each edge showed regular excursions around the mean (26), and a numerical diagnostic (27) monitoring the number edges in the sampled graph also clearly suggested that sufficient samples had been drawn ($\hat{R} \approx 1.0$).

The samples drawn using MCMC allow the posterior distribution of Bayesian networks to be approximated. In particular, the samples can be used to estimate the posterior edge probability $P(e|\mathbf{X})$ with $e \in E$. Figure 2 displays all edges with posterior probability of at least 0.5.

Our focus is on depression, the parents of which in Figure 2 we observe are "Didn't present to doctor" and "Gender". It important, however, to note that the model does not say that these are the only factors that are important. For example, "Drug user" at Wave I is related to depression through "Didn't present to doctor" at Wave I and II (Figure 2).

This is shown in Figure 3, which gives the conditional probability of being depressed at Wave 2 when a particular variable is set to a specific value. We see that general health, violence, academic performance and drug use all affect the conditional probability of depression at Wave II. Note that to compute this probability, links from the parents of the variable in which we 'intervene' are removed; this is equivalent to the 'do-operator' in the terminology of Pearl (28).

The analysis reveals the interaction between the many aspects of life that have an impact on depression. The connection between the depression and its two parents in Figure 2 have been previously discussed in the literature. The importance of gender in depression is particularly extensively documented in the literature (8). The connection to a failure in seeking medical care even when the individual thinks they should has also

been discussed in the literature, often in terms of poor accessibility of health care services for young people (29, 8). Several decades of research have revealed the complex causation of depression in young people, as suggested by this study (8).

4 DISCUSSION

There is a large amount of information held in large social science questionnaires. In this paper we have examined a graphical model approach to inferring structure amongst the variables in such questionnaires. In contrast to the standard regression-based approaches, a graphical model approach forgoes the need to specify a particular variable as the response. Instead, a more comprehensive estimate of the entire structure of the underlying system can be obtained. Regression approaches posit a particular conditional-independence structure, while graphical approaches allow consideration of more general structures.

The limitations of this study include those of all similar studies using observational data that are collected for multiple audiences. These forms of data, including the longitudinal data used here, do not permit strong causal conclusions to be drawn. In particular there may be important variables that we have not included in the analysis. However, the results are consistent with studies that have used other research approaches including experimental designs. The connection between an individual not seeking medical care when they think they should and depression supports current practice guidance in the UK (30) where there is an emphasis on providing access to health care through the school system rather than expecting young people to seek health care themselves. Not seeking medical care despite believing it should be sought is a complex factor because it captures both barriers to getting medical care within the individual, such as lacking motivation to seek care, and barriers within the individual's environment, such as poor access to care. This may mean that the variable encapsulates a number of different characteristics related to depression, and thus may form a 'marker' for depression. However, the use of a form of the question "Has there been any time over the past year when you thought you should get medical care, but you did not?" as a screening question in different contexts needs further consideration.

This method of analysis clarifies the complexity of depression and suggests why when using traditional methods of analysis it can be difficult to clarify whether or not factors, such as experiences in the family, in the wider community and at school, impact on the experience of depression for young people. It may also suggest why interventions for prevention of depression have not yet been demonstrated to be cost effective (31).

We performed structural inference for the Bayesian network using a Gibbs sampler (1), because MC^3 did not mix in a reasonable time. We have also found (1) this algorithm to be superior to the REV sampler (32), and it has the advantage of avoiding the need to consider an order prior as required by order MCMC methods (33, 34), which induces a bias that can only be corrected exactly by NP-hard computation of a correction factor.

An alternative to the MCMC method used here is the PC-algorithm (10, 11). This method is computationally efficient and is asymptotically consistent. However, to test whether the sample size available here is sufficient to reach the asymptotic regime, we applied the PC-algorithm (without constraints) to 10 different subsamples, each containing 90% of the data. We found that these results differed significantly, with a mean 84 in structural Hamming distance between the pairs of completed partially directed acyclic graphs (CPDAGs) given for the subsamples.

We used a Multinomial-Dirichlet model for the local conditional distributions, which yields a closed-form marginal likelihood. This model posits an entirely general discrete distribution, allowing its form to be guided by the data. However, the number of parameters in the local distributions for this model increases exponentially with the number of parents, which may mean that overly-sparse models are preferred. This is problematic when the sample size of the available data is small, because models with many parameters cannot be assessed adequately without a large dataset. The large sample size of the dataset used here minimises this issue, but it would nonetheless be worthwhile to consider more compact parameterisations. However, estimating such models (35) significantly increases the complexity of the model space, which makes such an approach computationally challenging in this setting.

For this paper, we removed samples with missing data. It is possible to handle missing data formally, for example by using structural EM (36), and similarly consider latent variables (e.g. shared genetics driving both child and parent behaviour). However, at present, doing so whilst robustly exploring large model spaces remains an open challenge. Tackling these computational and inferential issues is a key area for future research.

References

[1] Goudie, R. J. B. and Mukherjee, S. M. (2011). An Efficient Gibbs Sampler for Structural Inference in Bayesian Networks. CRiSM Working Paper 11-21 (Dept. of Statistics, University of Warwick).

[2] Friedman, N. (2004) *Science*, **303**, 5659, 799–805.

[3] Mukherjee, S. and Speed, T. P. (2008) *Proc Natl Acad Sci USA*, **105**, 38, 14313–14318.

[4] Acid, S., de Campos, L. M., Fernández-Luna, J. M., Rodríguez, S., Rodríguez, J. M. and Salcedo, J. L. (2004) *Artif Intell in Med*, **30**, 3, 215–232.

[5] Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G. and Angold, A. (2003) *Arch Gen Psych*, **60**, 8, 837–844.

[6] Costello, E. J., Erkanli, A. and Angold, A. (2006) *J Child Psychol Psych*, **47**, 12, 1263–1271.

[7] Thapar, A., Collishaw, S., Potter, R. and Thapar, A. K. (2010) *Br Med J*, **340**, c209.

[8] Patel, V., Flisher, A. J., Hetrick, S. and McGorry, P. (2007) *Lancet*, **369**, 9569, 1302–1313.

[9] Barnett, P. A. and Gotlib, I. H. (1988) *Psych Bull*, **104**, 1, 97–126.

[10] Spirtes, P., Glymour, C. and Scheines, R. (2000) *Causation, Prediction, and Search* (The MIT Press, Cambridge, MA).

[11] Korb, K. B. and Nicholson, A. E. (2011) *Bayesian Artificial Intelligence* (CRC Press, Boca Raton, FL).

[12] Harris, K. M., Halpern, C. T., Whitsel, E. A., Hussey, J., Tabor, J., Entzel, P. and Udry, J. R. (2009) The National Longitudinal Study of Adolescent Health: Research Design.

[13] Radloff, L. (1977) *App Psych Meas*, **1**, 3, 385–401.

[14] Goodman, E. (1999) *Am J Pub Health*, **89**, 10, 1522–1528.

[15] Roberts, R. E., Lewinsohn, P. M. and Seeley, J. R. (1991) *J Am Acad Child Adolesc Psych*, **30**, 1, 58–66.

[16] Holt, S., Buckley, H. and Whelan, S. (2008) *Child Abuse & Neglect*, **32**, 8, 797–810.

[17] Obot, I. S. and Anthony, J. C. (2004) *J Child Adolesc Subst Abuse*, **13**, 4, 83–96.

[18] Brown, R. A., Lewinsohn, P. M., Seeley, J. R. and Wagner, E. F. (1996) *J Am Acad Child Adolesc Psych*, **35**, 12, 1602–1610.

[19] Battles, H. B. and Wiener, L. S. (2002) *J Adoles Health*, **30**, 3, 161–168.

[20] Larun, L., Nordheim, L. V., Ekeland, E., Hagen, K. B. and Heian, F. (2006) *Cochrane Database Syst Rev*, 3, CD004691.

[21] Heckerman, D., Geiger, D. and Chickering, D. M. (1995) *Mach Learn*, **20**, 197–243.

[22] Madigan, D. and York, J. C. (1995) *Int Stat Rev*, **63**, 2, 215–232.

[23] Kalisch, M. and Bühlmann, P. (2007) *J Mach Learn Res*, **8**, 613–636.

[24] Cox, D. and Wermuth, N. (1996) *Multivariate Dependencies Models, Analysis and Interpretation* (Chapman & Hall, London).

[25] Robert, C. P. and Casella, G. (2004) *Monte Carlo Statistical Methods* (Springer, New York).

[26] Yu, B. and Mykland, P. (1998) *Statistics and Computing*, **8**, 3, 275–286.

[27] Gelman, A. and Rubin, D. B. (1992) *Statistical Science*, **7**, 4, 457–472.

[28] Pearl, J. (2000) *Causality: Models, Reasoning, and Inference* (Cambridge University Press, New York).

[29] Rickwood, D. J., Deane, F. P. and Wilson, C. J. (2007) *Med J Aus*, **187**, 7 Suppl, S35–S39.

[30] National Institute for Health and Clinical Excellence (2005) *Depression in Children and Young People* (NICE, London).

[31] Merry, S. N. (2007) *Curr Opin Psych*, **20**, 4, 325–329.

[32] Grzegorczyk, M. and Husmeier, D. (2008) *Mach Learn*, **71**, 2-3, 265–305.

[33] Ellis, B. and Wong, W. H. (2008) *J Am Stat Assoc*, **103**, 482, 778–789.

[34] Friedman, N. and Koller, D. (2003) *Mach Learn*, **50**, 1-2, 95–125.

[35] Friedman, N. and Goldszmidt, M. (1996) In *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96)* (Morgan Kaufmann Publishers Inc.), 252–260.

[36] Friedman, N. (1998) In *Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI-98)* (Morgan Kaufmann Publishers Inc.), 129–138.

Table 1: The table shows the label used in the plots above, the number of levels (r), and the exact wording of the question. The ID(s) of the relevant variables in the Add Health dataset are in parentheses. See www.cpc.unc.edu/projects/addhealth for full details of all of these questions.

Label	**r**	**Question**
Female	2	Interviewer, please confirm that R's sex is (male) female. (BIO_SEX)
Hisp/Latino	2	Are you of Hispanic or Latino origin? (H1GI4)
White	2	What is your race? [White] You may give more than one answer (H1GI6A)
Black/Af Am	2	What is your race? [Black or African American] You may give more than one answer (H1GI6B)
Am Ind/Nat Am	2	What is your race? [American Indian or Native American] You may give more than one answer (H1GI6C)
Asian/Pac Isl.	2	What is your race? [Asian or Pacific Islander] You may give more than one answer (H1GI6D)
Other race	2	What is your race? [Other] You may give more than one answer (H1GI6E)
Skips school	4	[If SCHOOL YEAR:] During this school year [If SUMMER:] During the 1994-1995 school year how many times HAVE YOU SKIPPED/DID YOU SKIP school for a full day without an excuse? (H1ED2; H2ED2)
Experiences prejudice	3	[If SCHOOL YEAR:] Students at your school are prejudiced [If SUMMER:] Last year, the students at your school were prejudiced. (H1ED21; H2ED17)
In physical fights	4	In the past 12 months, how often did you get into a serious physical fight? (H1DS5; H2FV16)
Didn't present to doctor	2	Has there been any time over the past year when you thought you should get medical care, but you did not? (H1GH26; H2GH28)
Severely injured	3	Which of these best describes your worst injury during the past year? (H1GH54; H2GH47)
Have HIV/AIDS	2	Have you ever been told by a doctor or a nurse that you had... HIV/AIDS (H1CO16D; H2CO19D)
Seen shooting	3	During the past 12 months, how often did each of the following things happen? You saw someone shoot or stab another person. (H1FV1; H2FV1)
Mother warm/loving	4	Most of the time, your mother is warm and loving toward you. (H1PF1; H2PF1)
Been suspended	2	Have you ever received an out-of-school suspension from school? (H1ED7; H2ED3)
Been expelled	2	Have you ever been expelled from school? (H1ED9; H2ED5)
Good health	3	In general, how is your health? Would you say... (H1GH1; H2GH1)
Talks to neighbours	2	In the past month, you have stopped on the street to talk with someone who lives in your neighborhood? (H1NB2; H2NB2)
Age	5	Age at interview, computed from date of birth, and date of interview (Constructed from IYEAR, IMONTH, IDAY, H1GI1Y, H1GI1M)
Live with mother	2	Indicator variable (Constructed from H1HR3A-T; H2HR4A-Q)
Live with father	2	Indicator variable (Constructed from H1HR3A-T; H2HR4A-Q)
Smoker	4	Frequency of smoking (Constructed from H1TO1/2/5; H2TO1/5)
Drinks alcohol	4	Frequency and amount of drinking alcohol (Constructed from H1TO12/15/18; H2TO15/19/22)
Exercises	3	Amount of exercise (Constructed from H1DA4/5/6; H2DA4-6)
Depressed	2	Rescaled CES-D, following (14) (Constructed from H1FS1-18; H2FS1-18)

Victim of violence	2	Indicator variable (Constructed from H1FV2-6; (H2FV2-5)
Family bereavement	3	Number of bereavements (Constructed from H1NM2/F2, H1FP24A1-5; H2NM4/F4, H2FP28A1-3)
Strong academically	4	Quartiles (Constructed from H1ED11-4; H2ED7-10)
Drug user	2	Indicator variable (Constructed from H1TO30/34/37/41; H2TO44/50/54/58)
Family poor	5	Census Bureau measure of poverty (Constructed from H1HR2/3/7/8, PA55)
Parents unhappy together	4	(Parent asked.) Do you and your partner argue/talk of separating? (Constructed from PB19/20)
Parent drinks	4	(Parent asked.) Number/frequency of drinks (Constructed from PA61/2)
Householder smokes	3	(Parent asked.) Either parent or others in household smokes (Constructed from PA63/4)
Has learning disability	2	(Parent asked.) Does (he/ she) have a specific learning disability, such as difficulties with attention, dyslexia, or some other reading, spelling, writing, or math disability? (PC38)
Parents aid decisions	5	(Parent asked.) How often would it be true for you to make each of the following statements about {child's name}? {Child's name} and you make decisions about (his/ her) life together. (PC34B)

Table 2: The groupings of the variables that were used to determine constraints on the Bayesian networks. Each variable in the analysis is either a Background variable, or from Wave I or Wave II of the Add Health study. Within each wave of the study, variables were further classified into whether they asked about the short- or long-term.

Background	Wave I Long-term	Wave I Short-term	Wave II Long-term	Wave II Short-term
Female	Skips school	Househol. smokes	Seen shooting	Smoker
Age	Experiences prejudice	Smoker	Alcohol	Live with mother
Hisp/Latino	In physical fights	Live with mother	Drug user	Live with father
White	Didn't pres. to doctor	Live with father	Mother warm/loving	Talks neighbours
Black/Af Am	Severely injured	Parent drinks	Have HIV/AIDS	Exercises
Am Ind/Nat Am	Have HIV/AIDS	Talks neighbours	Family bereavement	Depressed
Asian/Pac Isl.	Seen shooting	Exercises	Experiences prejudice	
Other race	Mother warm/loving	Depressed	Been expelled	
Has learning dis.	Been suspended		Been suspended	
	Been expelled		Victim of violence	
	Good health		In physical fights	
	Alcohol		Strong academically	
	Victim of violence		Didn't pres. to doctor	
	Family bereavement		Skips school	
	Strong academically		Severely injured	
	Drug user		Good health	
	Family poor			
	Parents unhappy togth.			
	Parents aid decisions			