## A VISL for missing-at-random scenario

In this section, we briefly explain why VISL can handle MAR problem by leveraging the results from Rubin [41]. Let's denote $\boldsymbol{r}_i$ as the missing mask, where $r_{i,d} = 1$ indicates $x_{i,d}$ is observed. For the random variable $R$, we use $p_\phi(\boldsymbol{r}|\mathbf{x})$ as the missing mechanism with parameters $\phi$. To explicitly state the dependence of VISL and its model parameter $\theta$, we use $p_\theta(\mathbf{x})$ to denote the corresponding model density. We can now formally define the concept of MAR (directly taken from [41]).

**Definition A.1** (Missing at Random [41])**.** The missing data are missing at random if for each value of $\phi$, $p_\phi(\boldsymbol{r}|\mathbf{x})$ takes the same value for all $\mathbf{x}_u$. Namely, $p_\phi(\boldsymbol{r}|\mathbf{x}) = p_\phi(\boldsymbol{r}|\mathbf{x}_o)$.

Recall that our VISL is trained by maximizing the ELBO (Eq.4) based on the observed values $x_o$. However, this formulation ignores the missing mechanisms $p_\phi(\boldsymbol{r}|\mathbf{x})$. In order to perform missing value imputation, one need to ensure that the inference for $\theta$ should be correct. In the following, we show under MAR, ignoring missing mechanism does not affect the correctness of inferring $\theta$ under ELBO. The following proof is an adaptation of Theorem 7.1 in Rubin [41].

When explicitly modelling the missing mechanism, the joint likelihood can be written as

$$\begin{aligned}
&\log p_{\theta,\phi}(\mathbf{x}_o, \boldsymbol{r})\\
&= \log \int p_\theta(\mathbf{x}) p_\phi(\boldsymbol{r}|\mathbf{x}) d\mathbf{x}_u\\
&= \log \int p_\theta(\mathbf{x}, \boldsymbol{z}, \boldsymbol{G}) p_\phi(\boldsymbol{r}|\mathbf{x}) d\boldsymbol{z} d\boldsymbol{G} d\mathbf{x}_u\\
&= \log p_\phi(\boldsymbol{r}|\mathbf{x}_o) \int p_\theta(\mathbf{x}, \boldsymbol{z}, \boldsymbol{G}) d\boldsymbol{z} d\boldsymbol{G} d\mathbf{x}_u\\
&\geq \log p_\phi(\boldsymbol{r}|\mathbf{x}_o) + \text{ELBO}(\theta)
\end{aligned}$$

where the third equality is from the definition of MAR and the last inequality is from the standard ELBO derivation. The above equation explicitly lower bounds the joint likelihood by two separate terms regarding $\theta$ and $\phi$. Thus, when performing inference over $\theta$, one can safely ignore the missing mechanism involving $\phi$, resulting in the same optimization objective as Eq.4. However, under a MNAR assumption, the missing mechanism $p_\phi(\mathbf{r}|\mathbf{x}_o, \mathbf{x}_u)$ can depend on the missing values. Therefore, $\log p_\phi(\mathbf{r}|\mathbf{x}_o, \mathbf{x}_u)$ cannot be separated out from the ELBO. Therefore, using the ELBO as a training objective will lead to a biased solution in a MNAR setting.

## B Does VISL respect the graph G in observational space

From the formulation of the decoder in VISL (Eq.5 and 6), the inferred graph $\mathbf{G}$ seems to define whether the information flow between nodes is allowed or not. Namely, when $G_{ij} = 1$, the information is allowed to pass from $z_i$ to $z_j$ at each iteration $t$. Thus, $\mathbf{G}$ directly defines a structure for latent space $\mathbf{Z}$, and indirectly defines a structure in observation $\mathbf{X}$ through the GNN updates and the final read-out layer. A natural question to ask is whether the resulting observations $\mathbf{x}$ from VISL also respect the graph $\mathbf{G}$. In the following, we show that when GNN is in equilibrium and the read-out layer is invertible without additional observational noise ($\sigma_x = 0$), the VISL is in fact a SEM for observation $\mathbf{x}$, which respects the graph $\mathbf{G}$. In the following, for the clarity of notations, we consider the structure learning between variables. For group-wise relations, it is trivial to generalize, since the going from variable-wise to group-wise only changes the read-out layer, where we use $M$ different MLPs instead of one.

First, let's clarify what do we mean by "respect a graph $\mathbf{G}$".

**Definition B.1** (Respect a graph $\mathbf{G}$)**.** For a given VISL model $p(\mathbf{x}, \mathbf{z}; \mathbf{G})$ with a specific graph $\mathbf{G}$, we say the model $p(\mathbf{x}; \mathbf{G}) = \int p(\mathbf{x}, \mathbf{z}; \mathbf{G}) d\mathbf{z}$ respects the graph $\mathbf{G}$ if it can be factorized

$$p(\mathbf{x}; \mathbf{G}) = \prod_{d=1}^{D} p(x_d|PA(d); \mathbf{G}),$$

where $PA(d)$ is a set of parents of node $i$ specified by graph $\mathbf{G}$.

### B.1 GNN at steady state

From the GNN message passing equations, we can re-organize Eq.5 and 6 into one equation:

$$z_i^t = F(PA(i)^{t-1}, z_i^{t-1}), \tag{11}$$

where $PA(i)^{t-1}$ is a set of parents' value for node $i$ at iteration $t-1$, $z_i^t$ is the value for node $i$ at iteration t and $F(\cdot)$ represents the GNN message passing updates.

The above equation resembles a fixed-point iteration procedure for function $F$. Indeed, under the context of GNN, this has been considered as a standard procedure to search for the equilibrium state due to its exponential convergence (Dai et al. 8, Eq.1; Gu et al. 16, Eq.2(b)). Thus, we assume that the GNN updates $F$ has unique equilibrium states given the initial conditions $AN(i)^0 \cup z_i^0$ for each $i$, where $AN(i)^0$ represents the initial values of ancestors for node $i$. For a sufficient condition of its existence, one can refer to Gu et al. [16, Theorem 4.1]. We note that this is only a sufficient condition, meaning that the GNN without the conditions in Gu et al. [16] can still have equilibrium states. Since discussing a necessary and sufficient conditions for the existence of the equilibrium state is out of the scope of this paper, we simply made an assumption that function $F$ has steady states. The reason we consider the initial ancestor values rather than just parent values is due to the message passing nature, where the value $PA(i)^t$ contains the information from the nodes that is at most $t$-hops away.

Since graph $G$ represents a DAG, one can always find a permutation $\pi$ of the original index $i = 1, \ldots, D$ based on the topological order. For concise notations, we assume the identity permutation. When the GNN is in equilibrium, we can rewrite Eq.11 as

$$z_i^\infty = F(PA^\infty(i), z_i^\infty), \tag{12}$$

where the superscript $\infty$ represents that steady state of the node. From the assumption, since the steady state for each $z_i^\infty$ depends on the initial values $AN(i)^0 \cup z_i^0$, it is trivial to see that the steady state $PA^\infty(i)$ depends on $AN^0(i)$. Therefore, the steady state $z_i^\infty$ is uniquely determined by $PA^\infty(i)$ and $z_i^0$. Namely,

$$z_i^\infty = H_i(PA^\infty(i), z_i^0) \tag{13}$$

for $i = 1, \ldots, D$, where $H_i$ is a mapping from $PA(i)^\infty$ and $z_i^0$ to the steady state of node $i$.

This is exactly the general form of an *structural equation model* (SEM) defined by graph $G$. If we further assume that the read-out layer $g(\cdot)$ is invertible, we can obtain

$$x_i = g\left(H_i\left(g^{-1}\left(PA_x^\infty(i)\right), z_i^0\right)\right), \tag{14}$$

which is also an SEM based on $G$ for observation $\mathbf{x}$. Thus, it is trivial that VISL respects the graph $G$ based on the above assumptions.

In practice, due to the exponential convergence of fixed point iteration, we found out that one does not need to use large iteration $t$. To balance the performance and computational cost, we found that 3 iterations of GNN message passing is enough to obtain reasonable performances.

## C Limitations

There are several exciting future research directions to address the potential limitations of VISL. First, although VISL is a structure learning method, we cannot claim that the inferred structure is causal. However, in Appendix B, we showed that when the GNN decoder reaches an equilibrium state with the invertible read-out layer, the learned joint distribution satisfies the causal Markov conditions, which is one of the fundamental assumptions for causal discovery. Thus, this opens a door for investigating what assumptions and modifications are required to make VISL a valid causal discovery approach.

Another limitation of VISL is the assumption about the type of missing mechanism. Currently, VISL can only handle MCAR or MAR but not MNAR. From Appendix A, MNAR mechanism can lead to biased model parameters and incorrect graph posterior.

Another potential future research direction is to design better variational distributions for the graph posterior and perform joint inference for both graph and functional parameters. Currently, the independent Bernoulli distribution cannot incorporate edge co-dependencies of the graph and the point estimate of the function parameters is not graph-dependent.

Under the current scope of this paper, we require the group $M$ is a predefined parameters for VISL. In real world, this information may not be available. Causal representation learning can be a candidate method to represent low level signals into groups. How to incorporate it with VISL is an interesting research question.

# D Experimental details

Here we specify the complete experimental details for full reproducibility. We first describe *Majority vote* and *Mean imputing* baselines, followed by all the details for the synthetic experiment. Then we explain the differences for the neuropathic pain and the Eedi topics experiments.

## D.1 Imputation baseline

**Majority vote**. *Majority vote* is a simple baseline method for missing value imputation. For an missing entry $x_{i,d}$, *Majority vote* will fill the gap by taking the most frequent value in $\mathbf{m}x_{:,d}$, i.e. the most frequent values in the column $d$.

**Mean imputing**. For a given missing entry $x_{i,d}$, *Mean imputing* will set $x_{i,d} = \frac{1}{N}\sum_{n=1}^{N} x_{n,d}$, i.e. the mean of column $d$.

## D.2 Synthetic experiment

**Data generation process**. To understand how the number of variables affects VISL, we use $D = 5, 7, 9$ variables (five datasets for each value of $D$). We first sample the underlying true structure. An edge from variable $i$ to variable $j$ is sampled with probability 0.5 if $i < j$, and probability 0 if $i \geq j$ (this ensures that the true structure is a DAG, which is just a standard scenario, and not a requirement for any of the compared algorithms). Then, we generate the data points. Root nodes (i.e. nodes with no parents, like variables 1 and 2 in Fig. 2(a) in the paper) are sampled from $\mathcal{N}(0, 1)$. Any other node $v_i$ is obtained from its parents $\mathrm{Pa}(i)$ as $v_i = \sum_{j \in \mathrm{Pa}(i)} \sin(3v_j) + \varepsilon$, where $\varepsilon \to \mathcal{N}(0, 0.01)$ is a Gaussian noise. We use the $\sin$ function to induce non-linear relationships between variables. Notice that the 3-times factor inside the $\sin$ encourages that the whole period of the $\sin$ function is used (to favor non-linearity). To evaluate the imputation methods, 30% of the test values are dropped. As an example of the data generation process, Fig. 7 below shows the pair plot for the dataset generated from the graph in Fig. 2(a) in the paper.

**Model parameters**. We start by specifying the parameters associated to the generative process. We use a prior probability $p_{ij} = 0.05$ in $\mathrm{p}(\mathbf{G})$ for all the edges. This favours sparse graphs, and can be adjusted depending on the problem at hand. The prior $\mathrm{p}(\mathbf{Z})$ is a standard Gaussian distribution, i.e. $\sigma_z^2 = 1$. This provides a standard regularisation for the latent space. The output noise is set to $\sigma_x^2 = 0.02$, which favours the accurate reconstruction of samples. As for the decoder, we perform $T = 3$ iterations of GNN message passing. All the MLPs in the decoder (i.e. $\text{MLP}^f$, $\text{MLP}^b$, $\text{MLP}^{e2n}$ and $g$) have two linear layers with ReLU non-linearity. The dimensionality of the hidden layer, which is the dimensionality of each latent subspace, is 256. Regarding the encoder, it is given by a multi-head neural network that defines the mean and standard deviation of the latent representation. The neural network is a MLP with two standard linear layers with ReLu non-linearity. The dimension of the hidden layer is also 256. When using groups, there are as many such MLPs as groups. Finally, recall that the variational posterior $\mathrm{q}(\mathbf{G})$ is the product of independent Bernoulli distributions over the edges, with a probability $\mathbf{G}_{ij}$ to be estimated for each edge. These values are all initialised to $\mathbf{G}_{ij} = 0.5$.

**Training hyperparameters**. We use Adam optimizer with learning rate $0.001$. We train during $300$ epochs with a batch size of $100$ samples. Each one of the two stages described in the two-step training takes half of the epochs. The percentage of data dropped during training for each instance is sampled from a uniform distribution. For the first half of training for structure learning, we adopt a procedure that gradually increases the dag penalty loss ($\lambda$ in Eq. 2). Initially, we set $\lambda = 150$ and $l = \lambda \mathbb{E}_{q(\mathbf{G})}[\mathcal{R}(\mathbf{G})]$. After every 50 epochs, we check whether the current dag loss $\lambda \mathbb{E}_{q(\mathbf{G})}[\mathcal{R}(\mathbf{G})] < 0.8l$. If not, we increase $\lambda = \lambda \times 8$, otherwise, $\lambda$ stays the same. Then, we update $l$ to be the currrent dag loss. This procedure is to make sure the final posterior graph is a DAG. When doing the reparametrization trick (i.e. when sampling from $\mathbf{Z}_n$), we obtain one sample during training

(100 samples in test time). For the Gumbel-softmax sample, we use a temperature $\tau = 0.5$. The rest of hyperparameters are the standard ones in `torch.nn.functional.gumbel_softmax`, in particular we use soft samples. The reason we use soft samples over hard samples is due to better empirical performances. Each soft sample is independently drawn from the Bernoulli variational graph posterior, and we then average the ELBO (Eq. 4) to compute the loss. To compute the DAG regulariser $\mathcal{R}(\mathbf{G})$, we use the exponential matrix implementation in `torch.matrix_exp`. This is in contrast to previous approaches, which resort to approximations [63, 61]. When applying the encoder, missing values in the training data are replaced with the value 0 (continuous variables). This trick is used to avoid dealing with variable sized inputs due to the different missingness for each entry $\mathbf{x}_i$. It can be regarded as a special imputation where we impute missing values as 0. One concern is that the formulation of MAR (Appx. A) and encoder $q_\phi(\mathbf{Z}|\tilde{\mathbf{x}}_i)$ (Eq. 9) requires the input $\tilde{\mathbf{x}}_i$ should only contain the observed information and no missing values should be used. In fact, this does not violate those formulations. Although the missing entry $x_{i,u}$ is replaced with value 0, this replacement is for all missing entries regardless of the position $d$ and index $i$. The value 0 should be treated as a constant rather than a variable, providing no additional information. Hence, the imputation $q_\phi(\mathbf{Z}|\tilde{\mathbf{x}}_i)$ still only uses the observed values. This trick has been known as *zero imputing* [30, 34]

**Baselines details**. Regarding the structure learning baselines, we ran both PC and GES with the Causal Command tool offered by the Center for Causal Discovery https://www.ccd.pitt.edu/tools/. We used the default parameters in each case (i.e. disc-bic-score for GES and cg-lr-test for PC). NOTEARS (L), NOTEARS (NL) and DAG-GNN were run with the code provided by the authors in GitHub: https://github.com/xunzheng/notears (NOTEARS (L) and NOTEARS (NL)) and https://github.com/fishmoon1234/DAG-GNN (DAG-GNN). In all cases, we used the default parameters proposed by the authors. Regarding the imputation baselines, Majority Vote and Mean Imputing were implemented in Python. MICE and Missforest were used from Scikit-learn library with default parameters https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html#sklearn.impute.IterativeImputer. For PVAE, we use the authors implementation with their proposed parameters, see https://github.com/microsoft/EDDI.

**Other experimental details**. VISL is implemented in PyTorch. The code is available in the supplementary material. The experiments were run using a local Tesla K80 GPU and a compute cluster provided by Azure Machine Learning platform with NVIDIA Tesla V100 GPU.

### D.3 Neuropathic pain experiment

**Data generation process**. We use the Neuropathic Pain Diagnosis Simulator in https://github.com/TURuibo/Neuropathic-Pain-Diagnosis-Simulator. We simulate five datasets with 1500 samples, and split each one randomly in 1000 training and 500 test samples. To evaluate the imputation methods, 30% of the test values are dropped. These five datasets are used for the five independent runs reported in experimental results.

**Model and training hyperparameters**. Most of the hyperparameters are identical to the synthetic experiment. However, in this case we have to deal with 222 variables, many more than before. In particular, the number of possible edges is 49062. Therefore, we reduce the dimensionality of each latent subspace to 32, the batch size to 25, and the amount of test samples for $\mathbf{Z}_n$ to 10 (in training we still use one as before). Moreover, we reduce the initial posterior probability for each edge to 0.2. The reason is that, for 0.5 initialization, the DAG regulariser $\mathcal{R}(\mathbf{G})$ evaluates to extremely high and unstable values for the $222 \times 222$ matrix. Since this is a more complex problem (no synthetic generation), we run the algorithm for 1000 epochs. When applying the encoder, missing values in the training data are replaced with the value 0.5 (binary variables).

### D.4 Eedi topics experiment

**Data generation process**. The real-world Eedi topics dataset contains 6147 samples, and can be downloaded from the website https://eedi.com/projects/neurips-education-challenge (task3_4 folder). The mapping from each question to its topics (also called "subjects") is given by the file "data/metadata/question_metadata_task_3_4.csv". For those questions that have more than one topic associated at the same level, randomly sample one of them. The hierarcy of topics (recall Fig. 4 in the paper) is given by the file "data/metadata/subject_metadata.csv". We use a random 80%-10%-10% train-validation-test split. The validation set is used to perform Bayesian Optimization (BO) as

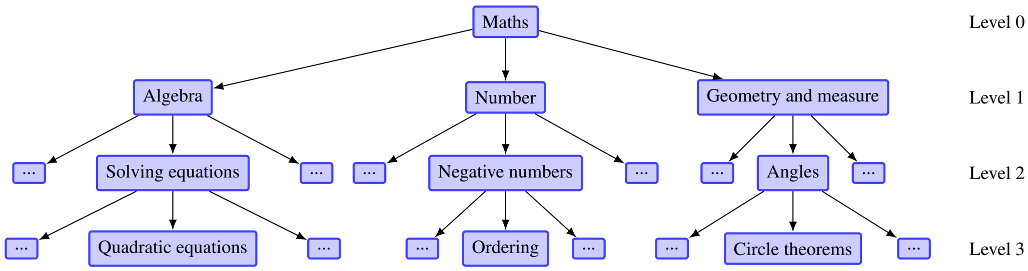

Figure 4: Hierarchy of topics in Eedi data. All the questions are related to maths (level 0 topic). The number of topics at levels 1, 2 and 3 are 3, 25 and 57. Each question is associated with only one topic at level 3 (thus, to only one topic at any higher level).

described below. The five runs reported in the experimental section come from different (random) initializations for the model parameters.

**Model and training hyperparameters**. Here, we follow the same specifications as in the neuropathic pain dataset. The only difference is that we perform BO for three hyperparameters: the dimensionality of the latent subspaces, the number of GNN message passing iterations, and the learning rate. The possible choices for each hyperparameter are $\{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$, $\{3, 5, 8, 10, 12, 14, 16, 18, 20\}$, and $\{10^{-4}, 10^{-3}, 10^{-2}\}$ respectively. We perform 39 runs of BO with the hyperdrive package in Azure Machine Learning platform https://docs.microsoft.com/en-us/python/api/azureml-train-core/azureml.train.hyperdrive?view=azure-ml-py. We use validation accuracy as the target metric. The best configuration obtained through BO was $15$, $8$ and $10^{-4}$, respectively.

**Baselines details**. As explained in the paper, in this experiment DAG-GNN is adapted to deal with missing values and groups of arbitrary size. For the former, we adapt the DAG-GNN code to replace missing values with $0.5$ constant value, as in VISL. For the latter, we also follow VISL and use as many different neural networks as groups (as described in the paper), all of them with the same architecture as the one used in the original code (https://github.com/fishmoon1234/DAG-GNN).

**Other experimental details**. The list of relationships found by VISL (Table 15) and DAG-GNN (Table 16) aggregates the relationships obtained in the five independent runs. This is done by setting a threshold of $0.35$ on the posterior probability of edge (which is initialized to $0.2$) and considering the union for the different runs. This resulted in $50$ relationships for VISL and $57$ for DAG-GNN. For *Random*, we simulated $50$ random relationships. Also, the probability reported in the first column of Table 15 is the average of the probabilities obtained for that relationship in the five different runs.

# E Additional figures, tables and results

| VISL | Number | Algebra | Geometry |
|---|---|---|---|
| Number | 30 | 4 | 3 |
| Algebra | 2 | 6 | 0 |
| Geometry | 0 | 0 | 5 |

| DAG-GNN | Number | Algebra | Geometry |
|---|---|---|---|
| Number | 8 | 3 | 6 |
| Algebra | 1 | 5 | 2 |
| Geometry | 14 | 7 | 11 |

| *Random* | Number | Algebra | Geometry |
|---|---|---|---|
| Number | 7 | 4 | 6 |
| Algebra | 8 | 1 | 6 |
| Geometry | 6 | 3 | 9 |

Table 7: Distribution of the relationships across level 1 topics.The item $(i, j)$ refers to edges $i \to j$.

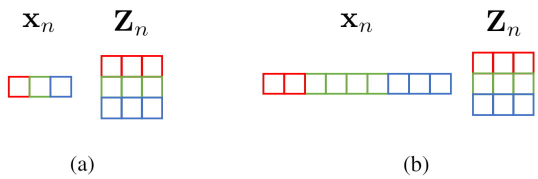


Figure 5: Structured latent space. (a) At the level of variables. Each variable in $\mathbf{x}_n$ (each color) has its own latent subspace, which is given by a row in $\mathbf{Z}_n$. (b) At the level of groups of variables. Here, each group of variables (each color) has its own latent subspace, which is given by a row in $\mathbf{Z}_n$.

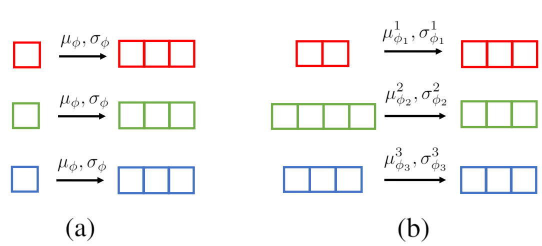


Figure 6: The encoder respects the structure of the latent space. (a) At the level of variables. All the variables use the same encoding functions. (b) At the level of groups of variables. Each group of variables uses different encoding functions.

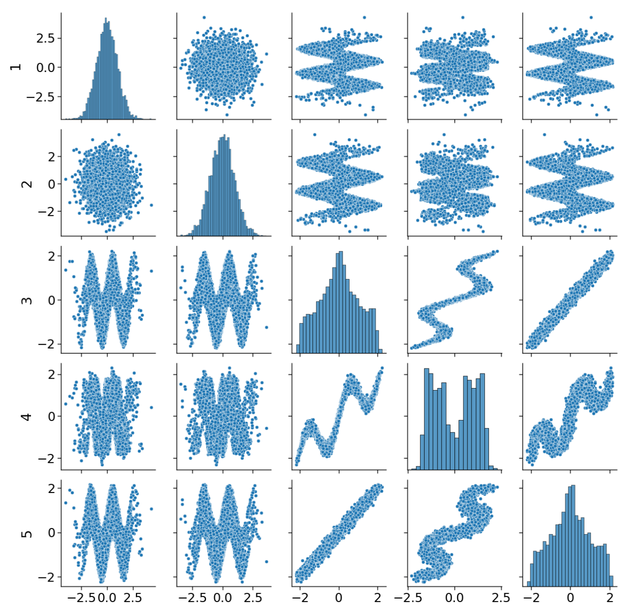


Figure 7: Pair-plot for the dataset generated from the graph in Fig. 2(a) in the paper. We observe different type of relationships between variables, including non-linear ones.

| | Number of variables | | | Average |
|---|---|---|---|---|
| | 5 | 7 | 9 | |
| Majority vote | 0.5507±0.0056 | 0.5391±0.0050 | 0.5427±0.0050 | 0.5442±0.0032 |
| Mean imputing | 0.2351±0.0104 | 0.2124±0.0112 | 0.2143±0.0064 | 0.2206±0.0061 |
| MICE | 0.1352±0.0044 | 0.1501±0.0095 | 0.1230±0.0025 | 0.1361±0.0046 |
| Missforest | 0.1279±0.0040 | 0.1403±0.0030 | 0.1258±0.0022 | 0.1313±0.0025 |
| PVAE | 0.1324±0.0048 | 0.1536±0.0095 | 0.1360±0.0019 | 0.1407±0.0043 |
| VISL | **0.1122±0.0029** | **0.1282±0.0037** | **0.1202±0.0006** | **0.1202±0.0040** |

Table 8: Imputation results for the synthetic experiment in terms of RMSE (not aggregating by number of variables, $D = 5, 7, 9$). The values are the mean and standard error over five different simulations.

| Index | Topic name |
|---|---|
| 1 | Decimals |
| 2 | Factors, Multiples and Primes |
| 3 | Fractions, Decimals and Percentage Equivalence |
| 4 | Fractions |
| 5 | Indices, Powers and Roots |
| 6 | Negative Numbers |
| 7 | Straight Line Graphs |
| 8 | Inequalities |
| 9 | Sequences |
| 10 | Writing and Simplifying Expressions |
| 11 | Angles |
| 12 | Circles |
| 13 | Co-ordinates |
| 14 | Construction, Loci and Scale Drawing |
| 15 | Symmetry |
| 16 | Units of Measurement |
| 17 | Volume and Surface Area |
| 18 | Basic Arithmetic |
| 19 | Factorising |
| 20 | Solving Equations |
| 21 | Formula |
| 22 | 2D Names and Properties of Shapes |
| 23 | Perimeter and Area |
| 24 | Similarity and Congruency |
| 25 | Transformations |

Table 9: Mapping between indexes for row/column names in Table 12 and Table 14 and the actual level-2 topic names.

| | | Adjacency | | | Orientation | | | Causal Accuracy |
|---|---|---|---|---|---|---|---|---|
| | | Recall | Precision | $F_1$-score | Recall | Precision | $F_1$-score | |
| 5 | PC | 0.464±0.099 | 0.610±0.117 | 0.526±0.107 | 0.364±0.098 | 0.490±0.127 | 0.416±0.111 | 0.436±0.076 |
| | GES | 0.414±0.067 | 0.507±0.071 | 0.446±0.065 | 0.257±0.103 | 0.327±0.117 | 0.285±0.110 | 0.368±0.072 |
| | NOTEARS (L) | 0.186±0.052 | 0.400±0.089 | 0.247±0.063 | 0.119±0.049 | 0.300±0.110 | 0.167±0.065 | 0.119±0.049 |
| | NOTEARS (NL) | 0.331±0.057 | 0.470±0.078 | 0.384±0.065 | 0.264±0.047 | 0.370±0.053 | 0.304±0.049 | 0.264±0.047 |
| | DAG-GNN | 0.381±0.130 | 0.433±0.121 | 0.399±0.127 | 0.231±0.067 | 0.283±0.073 | 0.249±0.068 | 0.231±0.067 |
| | VISL | 0.943±0.051 | 0.708±0.092 | 0.779±0.050 | 0.943±0.051 | 0.708±0.092 | 0.779±0.050 | 0.943±0.026 |
| 7 | PC | 0.396±0.110 | 0.639±0.154 | 0.468±0.112 | 0.113±0.043 | 0.193±0.083 | 0.134±0.050 | 0.324±0.088 |
| | GES | 0.429±0.087 | 0.647±0.042 | 0.501±0.076 | 0.208±0.067 | 0.279±0.081 | 0.235±0.073 | 0.345±0.091 |
| | NOTEARS (L) | 0.222±0.059 | 0.526±0.124 | 0.309±0.078 | 0.176±0.041 | 0.436±0.109 | 0.248±0.058 | 0.176±0.041 |
| | NOTEARS (NL) | 0.315±0.094 | 0.513±0.119 | 0.382±0.104 | 0.269±0.074 | 0.453±0.105 | 0.330±0.084 | 0.269±0.074 |
| | DAG-GNN | 0.396±0.109 | 0.539±0.123 | 0.446±0.111 | 0.318±0.082 | 0.445±0.102 | 0.361±0.085 | 0.318±0.082 |
| | VISL | 0.764±0.078 | 0.855±0.048 | 0.8±0.061 | 0.636±0.103 | 0.702±0.087 | 0.717±0.069 | 0.636±0.103 |
| 9 | PC | 0.406±0.072 | 0.654±0.053 | 0.491±0.060 | 0.176±0.020 | 0.302±0.045 | 0.219±0.024 | 0.229±0.041 |
| | GES | 0.514±0.065 | 0.553±0.050 | 0.525±0.049 | 0.282±0.057 | 0.308±0.068 | 0.291±0.061 | 0.379±0.069 |
| | NOTEARS (L) | 0.172±0.026 | 0.403±0.076 | 0.238±0.036 | 0.151±0.023 | 0.366±0.082 | 0.211±0.035 | 0.151±0.023 |
| | NOTEARS (NL) | 0.338±0.042 | 0.485±0.053 | 0.394±0.045 | 0.297±0.034 | 0.429±0.044 | 0.347±0.036 | 0.297±0.034 |
| | DAG-GNN | 0.551±0.067 | 0.554±0.053 | 0.547±0.057 | 0.508±0.061 | 0.516±0.054 | 0.508±0.055 | 0.508±0.061 |
| | VISL | 0.623±0.048 | 0.640±0.051 | 0.626±0.045 | 0.389±0.076 | 0.403±0.078 | 0.398±0.077 | 0.398±0.076 |

Table 10: Structure learning results for the synthetic experiment (not aggregating by number of variables, $D = 5, 7, 9$). The values are the mean and standard error over five different simulations.

| | Number of variables | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 |
| PC | 2.49±0.62 | 5.19±1.01 | 8.14±1.64 | 14.99±1.59 | 21.65±2.19 | 26.11±1.70 | 30.21±2.01 | 35.43±1.56 |
| GES | 0.21±0.02 | 1.12±0.41 | 1.80±0.80 | 2.28±0.78 | 2.76±1.01 | 3.34±0.52 | 3.87±0.66 | 4.10±0.71 |
| NOTEARS (L) | 8.91±2.34 | 21.04±3.43 | 38.53±2.52 | 56.11±3.23 | 91.11±4.15 | 140.33±3.53 | 331.04±6.55 | 378.21±9.12 |
| NOTEARS (NL) | 12.94±2.18 | 31.03±3.11 | 54.08±4.10 | 89.35±4.11 | 99.32±5.12 | 240.43±5.39 | 364.92±3.22 | 469.43±4.77 |
| DAG-GNN | 13.62±2.93 | 30.04±2.48 | 52.01±3.81 | 88.12±79 | 112.33±5.01 | 255.11±6.93 | 371.22±5.32 | 498.09±5.01 |
| VISL | 10.27±1.98 | 25.11±5.21 | 47.98±3.12 | 76.12±4.40 | 101.12±4.23 | 201.59±6.33 | 340.10±8.22 | 421.11±5.33 |
| MMHC | 7.85±1.02 | 69.10±5.32 | 542.92±9.82 | 1314.76±9.10 | NA | NA | NA | NA |
| Tabu | 2.01±0.72 | 7.45±1.03 | 24.08±5.93 | 57.87±3.85 | 77.87±5.52 | 128.67±4.09 | 163.33±6.55 | 219.05±3.42 |
| HillClimb | 1.52±0.64 | 6.98±1.23 | 22.10±5.32 | 51.78±4.06 | 75.29±5.84 | 121.92±4.71 | 157.82±6.87 | 209.54±5.01 |

Table 11: Running times (in minutes) for different structure learning approaches in an extended synthetic experiment. For each number of variables, three datasets were simulated following the same data generation process described above, and the results show the mean and standard error. Notice that we have considered three additional baselines (MMHC, Tabu, HillClimb). We observe three different types of methods. MMHC scales poorly (NA means that the training took more than 24 hours), probably due to its hybrid nature that combines constraint-based and score-based approaches. VISL and the other deep learning based methods (DAG-GNN, NOTEARS) can scale to large numbers of variables. Of course, simpler methods such as PC, GES, Tabu, and HillClimb are significantly faster than VISL (note that these baselines are from highly optimized libraries that leverage e.g. dynamic programming and parallelization), but their performance is worse. Indeed, the structure learning performance for this experiment is shown in Table 18

.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 5 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 12: How the 50 relationships found by VISL are distributed across level 2 topics. The item $(i, j)$ refers to edges in the direction $i \rightarrow j$. There are 18 relationships inside level 2 topics (36%). See Table 9 for a mapping between indexes shown here in row/column names and the actual level-2 topic names.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| 14 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |

Table 13: How the 57 relationships found by DAG-GNN are distributed across level 2 topics. The item $(i, j)$ refers to edges in the direction $i \rightarrow j$. There are 8 relationships inside level 2 topics (14%). See Table 9 for a mapping between indexes shown here in row/column names and the actual level-2 topic names.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0 | 3 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 16 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 19 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 24 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 14: How the 50 relationships found by *Random* are distributed across level 2 topics. The item $(i, j)$ refers to edges in the direction $i \rightarrow j$. There are 3 relationships inside level 2 topics (6%). See Table 9 for a mapping between indexes shown here in row/column names and the actual level-2 topic names.

| Prob | Topic 1 (from) | Topic 2 (to) | Adj1 | Ori1 | Adj2 | Ori2 |
|---|---|---|---|---|---|---|
| 0.44 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Ordering Negative Numbers [Negative Numbers] [Number] | 5 | 1 | 5 | 1 |
| 0.38 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 5 | 5 | 5 | 5 |
| 0.37 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 5 | 5 | 5 | 5 |
| 0.37 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 2 | 2 | 2 | 1 |
| 0.36 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 2 | 1 | 2 | 1 |
| 0.35 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Place Value [Basic Arithmetic] [Number] | 4 | 2 | 4 | 2 |
| 0.35 | Mental Multiplication and Division [Basic Arithmetic] [Number] | BIDMAS [Basic Arithmetic] [Number] | 5 | 5 | 5 | 5 |
| 0.35 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 5 | 5 | 5 | 5 |
| 0.35 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | BIDMAS [Basic Arithmetic] [Number] | 4 | 4 | 4 | 3 |
| 0.35 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | 4 | 4 | 5 | 4 |
| 0.35 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Squares, Cubes, etc [Indices, Powers and Roots] [Number] | 5 | 5 | 5 | 5 |
| 0.34 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 5 | 1 | 5 | 1 |
| 0.34 | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | Angle Facts with Parallel Lines [Angles] [Geometry and Measure] | 4 | 4 | 4 | 4 |
| 0.34 | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 4 | 2 | 5 | 2 |
| 0.34 | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 5 | 2 | 5 | 2 |
| 0.34 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Squares, Cubes, etc [Indices, Powers and Roots] [Number] | 2 | 2 | 2 | 2 |
| 0.33 | Ordering Negative Numbers [Negative Numbers] [Number] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 5 | 5 | 5 | 5 |
| 0.33 | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | Measuring Angles [Angles] [Geometry and Measure] | 3 | 2 | 5 | 2 |
| 0.33 | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | 4 | 4 | 4 | 4 |
| 0.33 | Measuring Angles [Angles] [Geometry and Measure] | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | 3 | 3 | 5 | 3 |
| 0.33 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Place Value [Basic Arithmetic] [Number] | 4 | 1 | 4 | 1 |
| 0.33 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 2 | 2 | 2 | 1 |
| 0.33 | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | BIDMAS [Basic Arithmetic] [Number] | 4 | 4 | 4 | 4 |
| 0.32 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | BIDMAS [Basic Arithmetic] [Number] | 3 | 2 | 3 | 2 |
| 0.32 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 5 | 5 | 5 | 5 |
| 0.32 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 2 | 1 | 2 | 1 |
| 0.32 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 3 | 3 | 3 | 3 |
| 0.32 | Linear Equations [Solving Equations] [Algebra] | Substitution into Formula [Formula] [Algebra] | 4 | 2 | 4 | 2 |
| 0.32 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Squares, Cubes, etc [Indices, Powers and Roots] [Number] | 3 | 2 | 3 | 2 |
| 0.32 | Angle Facts with Parallel Lines [Angles] [Geometry and Measure] | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | 4 | 2 | 4 | 2 |
| 0.32 | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Substitution into Formula [Formula] [Algebra] | 2 | 2 | 2 | 2 |
| 0.32 | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | Substitution into Formula [Formula] [Algebra] | 4 | 3 | 4 | 3 |
| 0.32 | Mental Multiplication and Division [Basic Arithmetic] [Number] | Time [Units of Measurement] [Geometry and Measure] | 4 | 4 | 4 | 4 |
| 0.32 | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Ordering Negative Numbers [Negative Numbers] [Number] | 4 | 2 | 4 | 2 |
| 0.32 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Substitution into Formula [Formula] [Algebra] | 5 | 5 | 5 | 5 |
| 0.32 | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 2 | 1 | 2 | 1 |
| 0.31 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 5 | 5 | 5 | 5 |
| 0.31 | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | Types, Naming and Estimating [Angles] [Geometry and Measure] | 4 | 2 | 5 | 2 |
| 0.31 | Ordering Negative Numbers [Negative Numbers] [Number] | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | 4 | 4 | 4 | 4 |
| 0.31 | Substitution into Formula [Formula] [Algebra] | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | 4 | 3 | 4 | 3 |
| 0.31 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | 2 | 2 | 2 | 1 |
| 0.31 | BIDMAS [Basic Arithmetic] [Number] | Place Value [Basic Arithmetic] [Number] | 4 | 2 | 4 | 1 |
| 0.31 | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 4 | 2 | 4 | 2 |
| 0.31 | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 4 | 2 | 4 | 2 |
| 0.30 | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | 2 | 2 | 1 | 1 |
| 0.30 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Time [Units of Measurement] [Geometry and Measure] | 2 | 2 | 2 | 2 |
| 0.30 | Ordering Negative Numbers [Negative Numbers] [Number] | Substitution into Formula [Formula] [Algebra] | 3 | 3 | 2 | 2 |
| 0.30 | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Angles in Polygons [Angles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| 0.30 | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Place Value [Basic Arithmetic] [Number] | 3 | 2 | 3 | 1 |
| 0.28 | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 2 | 1 | 2 | 1 |

Table 15: Full list of relationships found by VISL in the Eedi topics dataset. Each row refers to one relationship (one edge). From left to right, the columns are the posterior probability of the edge, the sending node (topic), the receiving node (topic), and the adjacency and orientation evaluations from each expert. For each topic, the brackets contain its parent level 2 and level 1 topics.

| Topic 1 (From) | Topic 2 (To) | Adj1 | Ori1 | Adj2 | Ori2 |
|---|---|---|---|---|---|
| Missing Lengths [Perimeter and Area] [Geometry and Measure] | Midpoint Between Two Co-ordinates [Co-ordinates] [Algebra] | 4 | 4 | 4 | 5 |
| Construct Triangle [Construction, Loci and Scale Drawing] [Geometry and Measure] | Place Value [Basic Arithmetic] [Number] | 1 | 1 | 1 | 1 |
| Squares, Cubes, etc [Indices, Powers and Roots] [Number] | Volume of Prisms [Volume and Surface Area] [Geometry and Measure] | 4 | 5 | 5 | 4 |
| Converting between Fractions and Percentages [Fractions, Decimals and Percentage Equivalence] [Number] | Volume of Prisms [Volume and Surface Area] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Angles in Triangles [Angles] [Geometry and Measure] | Parts of a Circle [Circles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Angle Facts with Parallel Lines [Angles] [Geometry and Measure] | 4 | 5 | 5 | 5 |
| Mental Multiplication and Division [Basic Arithmetic] [Number] | Measuring Angles [Angles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Angles in Polygons [Angles] [Geometry and Measure] | Compound Area [Perimeter and Area] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Squares, Cubes, etc [Indices, Powers and Roots] [Number] | Solving Linear Inequalities [Inequalities] [Algebra] | 2 | 1 | 3 | 1 |
| Construct Triangle [Construction, Loci and Scale Drawing] [Geometry and Measure] | Solving Linear Inequalities [Inequalities] [Algebra] | 1 | 1 | 1 | 1 |
| Written Multiplication [Basic Arithmetic] [Number] | Translation and Vectors [Transformations] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Enlargement [Transformations] [Geometry and Measure] | Reflection [Transformations] [Geometry and Measure] | 5 | 2 | 5 | 3 |
| Rotation [Transformations] [Geometry and Measure] | Reflection [Transformations] [Geometry and Measure] | 4 | 3 | 5 | 2 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Length Scale Factors in Similar Shapes [Similarity and Congruency] [Geometry and Measure] | 1 | 1 | 2 | 1 |
| Angles in Triangles [Angles] [Geometry and Measure] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 4 | 3 | 5 | 3 |
| Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 1 | 1 | 3 | 1 |
| Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Written Multiplication [Basic Arithmetic] [Number] | Perimeter [Perimeter and Area] [Geometry and Measure] | 2 | 1 | 2 | 1 |
| Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | Perimeter [Perimeter and Area] [Geometry and Measure] | 2 | 1 | 4 | 1 |
| Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | Area of Simple Shapes [Perimeter and Area] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Substitution into Formula [Formula] [Algebra] | Writing Expressions [Writing and Simplifying Expressions] [Algebra] | 4 | 2 | 3 | 1 |
| Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | Linear Equations [Solving Equations] [Algebra] | 1 | 1 | 1 | 1 |
| Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | Factorising into a Single Bracket [Factorising] [Algebra] | 4 | 5 | 5 | 4 |
| Linear Equations [Solving Equations] [Algebra] | Factorising into a Single Bracket [Factorising] [Algebra] | 4 | 3 | 5 | 3 |
| Converting between Fractions and Decimals [Fractions, Decimals and Percentage Equivalence] [Number] | BIDMAS [Basic Arithmetic] [Number] | 1 | 1 | 1 | 1 |
| Reflection [Transformations] [Geometry and Measure] | Place Value [Basic Arithmetic] [Number] | 1 | 1 | 1 | 1 |
| Length, Area and Volume Scale Factors [Similarity and Congruency] [Geometry and Measure] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 5 | 1 | 4 | 1 |
| Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | Midpoint Between Two Co-ordinates [Co-ordinates] [Algebra] | 5 | 5 | 5 | 5 |
| Enlargement [Transformations] [Geometry and Measure] | Time [Units of Measurement] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Rotational Symmetry [Symmetry] [Geometry and Measure] | Midpoint Between Two Co-ordinates [Co-ordinates] [Algebra] | 1 | 1 | 2 | 1 |
| Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | Horizontal and Vertical Lines [Straight Line Graphs] [Algebra] | 1 | 1 | 1 | 1 |
| Angles in Triangles [Angles] [Geometry and Measure] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Rotational Symmetry [Symmetry] [Geometry and Measure] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Equivalent Fractions [Fractions] [Number] | Converting Mixed Number and Improper Fractions [Fractions] [Number] | 5 | 5 | 5 | 5 |
| Multiplying and Dividing with Decimals [Decimals] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Construct Angle [Construction, Loci and Scale Drawing] [Geometry and Measure] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Angle Facts with Parallel Lines [Angles] [Geometry and Measure] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Measuring Angles [Angles] [Geometry and Measure] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 5 | 1 | 4 | 5 |
| Squares, Cubes, etc [Indices, Powers and Roots] [Number] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 3 | 1 |
| Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 5 | 2 | 5 | 1 |
| Ordering Negative Numbers [Negative Numbers] [Number] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 5 | 5 | 5 | 5 |
| Rotation [Transformations] [Geometry and Measure] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 3 | 1 |
| Reflection [Transformations] [Geometry and Measure] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 3 | 1 |
| Perimeter [Perimeter and Area] [Geometry and Measure] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Converting between Fractions and Percentages [Fractions, Decimals and Percentage Equivalence] [Number] | Ordering Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Ordering Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Perimeter [Perimeter and Area] [Geometry and Measure] | Ordering Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Time [Units of Measurement] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Written Multiplication [Basic Arithmetic] [Number] | BIDMAS [Basic Arithmetic] [Number] | 5 | 4 | 5 | 3 |

Table 16: Full list of relationships found by DAG-GNN in the Eedi topics dataset. Each row refers to one relationship (one edge). From left to right, the columns are the sending node (topic), the receiving node (topic), and the adjacency and orientation evaluations from each expert. For each topic, the brackets contain its parent level 2 and level 1 topics.

| Topic 1 (From) | Topic 2 (To) | Adj1 | Ori1 | Adj2 | Ori2 |
|---|---|---|---|---|---|
| Midpoint Between Two Co-ordinates [Co-ordinates] [Algebra] | Angles in Triangles [Angles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Solving Linear Inequalities [Inequalities] [Algebra] | Enlargement [Transformations] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Squares, Cubes, etc [Indices, Powers and Roots] [Number] | Written Multiplication [Basic Arithmetic] [Number] | 4 | 1 | 5 | 1 |
| Substitution into Formula [Formula] [Algebra] | Written Multiplication [Basic Arithmetic] [Number] | 4 | 1 | 3 | 1 |
| Linear Sequences (nth term) [Sequences] [Algebra] | Mental Multiplication and Division [Basic Arithmetic] [Number] | 5 | 1 | 5 | 2 |
| Measuring Angles [Angles] [Geometry and Measure] | Construct Angle [Construction, Loci and Scale Drawing] [Geometry and Measure] | 5 | 5 | 5 | 5 |
| Dividing Fractions [Fractions] [Number] | Volume of Prisms [Volume and Surface Area] [Geometry and Measure] | 2 | 2 | 2 | 2 |
| Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | Parts of a Circle [Circles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Types, Naming and Estimating [Angles] [Geometry and Measure] | Parts of a Circle [Circles] [Geometry and Measure] | 2 | 2 | 2 | 1 |
| Angles in Polygons [Angles] [Geometry and Measure] | Basic Angle Facts (straight line, opposite, around a point, etc) [Angles] [Geometry and Measure] | 5 | 1 | 5 | 1 |
| Angles in Polygons [Angles] [Geometry and Measure] | Compound Area [Perimeter and Area] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Length, Area and Volume Scale Factors [Similarity and Congruency] [Geometry and Measure] | Linear Sequences (nth term) [Sequences] [Algebra] | 1 | 1 | 2 | 1 |
| Substitution into Formula [Formula] [Algebra] | Rotation [Transformations] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Gradient Between Two Co-ordinates [Co-ordinates] [Algebra] | 5 | 5 | 5 | 5 |
| Compound Area [Perimeter and Area] [Geometry and Measure] | Reflection [Transformations] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| BIDMAS [Basic Arithmetic] [Number] | Reflection [Transformations] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Compound Area [Perimeter and Area] [Geometry and Measure] | Properties of Quadrilaterals [2D Names and Properties of Shapes] [Geometry and Measure] | 3 | 1 | 3 | 1 |
| Rotational Symmetry [Symmetry] [Geometry and Measure] | Perimeter [Perimeter and Area] [Geometry and Measure] | 3 | 1 | 3 | 1 |
| Converting between Fractions and Percentages [Fractions, Decimals and Percentage Equivalence] [Number] | Area of Simple Shapes [Perimeter and Area] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Angles in Triangles [Angles] [Geometry and Measure] | Types, Naming and Estimating [Angles] [Geometry and Measure] | 4 | 3 | 5 | 2 |
| Length Scale Factors in Similar Shapes [Similarity and Congruency] [Geometry and Measure] | Types, Naming and Estimating [Angles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Factorising into a Single Bracket [Factorising] [Algebra] | Types, Naming and Estimating [Angles] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Enlargement [Transformations] [Geometry and Measure] | BIDMAS [Basic Arithmetic] [Number] | 1 | 1 | 1 | 1 |
| Linear Sequences (nth term) [Sequences] [Algebra] | Time [Units of Measurement] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Horizontal and Vertical Lines [Straight Line Graphs] [Algebra] | Adding and Subtracting Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Area of Simple Shapes [Perimeter and Area] [Geometry and Measure] | Multiplying and Dividing Negative Numbers [Negative Numbers] [Number] | 1 | 1 | 1 | 1 |
| Writing Expressions [Writing and Simplifying Expressions] [Algebra] | Factors and Highest Common Factor [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Squares, Cubes, etc [Indices, Powers and Roots] [Number] | Midpoint Between Two Co-ordinates [Co-ordinates] [Algebra] | 1 | 1 | 1 | 1 |
| Writing Expressions [Writing and Simplifying Expressions] [Algebra] | Naming Co-ordinates in 2D [Co-ordinates] [Algebra] | 1 | 1 | 1 | 1 |
| BIDMAS [Basic Arithmetic] [Number] | Line Symmetry [Symmetry] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Length, Area and Volume Scale Factors [Similarity and Congruency] [Geometry and Measure] | 1 | 1 | 1 | 1 |
| Converting Mixed Number and Improper Fractions [Fractions] [Number] | Horizontal and Vertical Lines [Straight Line Graphs] [Algebra] | 1 | 1 | 1 | 1 |
| Construct Angle and Line Bisectors [Construction, Loci and Scale Drawing] [Geometry and Measure] | Horizontal and Vertical Lines [Straight Line Graphs] [Algebra] | 1 | 1 | 1 | 1 |
| Multiplying and Dividing with Decimals [Decimals] [Number] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Reflection [Transformations] [Geometry and Measure] | Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | 1 | 1 | 1 | 1 |
| Substitution into Formula [Formula] [Algebra] | Dividing Fractions [Fractions] [Number] | 4 | 1 | 3 | 1 |
| Factorising into a Single Bracket [Factorising] [Algebra] | Dividing Fractions [Fractions] [Number] | 2 | 1 | 2 | 1 |
| Fractions of an Amount [Fractions] [Number] | Multiplying Fractions [Fractions] [Number] | 5 | 4 | 5 | 2 |
| Time [Units of Measurement] [Geometry and Measure] | Converting Mixed Number and Improper Fractions [Fractions] [Number] | 4 | 1 | 4 | 1 |
| Length Scale Factors in Similar Shapes [Similarity and Congruency] [Geometry and Measure] | Converting Mixed Number and Improper Fractions [Fractions] [Number] | 4 | 1 | 5 | 1 |
| Place Value [Basic Arithmetic] [Number] | Equivalent Fractions [Fractions] [Number] | 4 | 4 | 3 | 5 |
| Reflection [Transformations] [Geometry and Measure] | Equivalent Fractions [Fractions] [Number] | 1 | 1 | 1 | 1 |
| Writing Expressions [Writing and Simplifying Expressions] [Algebra] | Fractions of an Amount [Fractions] [Number] | 1 | 1 | 1 | 1 |
| Dividing Fractions [Fractions] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Adding and Subtracting Fractions [Fractions] [Number] | Prime Numbers and Prime Factors [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Simplifying Expressions by Collecting Like Terms [Writing and Simplifying Expressions] [Algebra] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 1 | 1 | 1 | 1 |
| Adding and Subtracting Fractions [Fractions] [Number] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 1 | 1 | 2 | 1 |
| Mental Multiplication and Division [Basic Arithmetic] [Number] | Multiples and Lowest Common Multiple [Factors, Multiples and Primes] [Number] | 5 | 5 | 5 | 5 |
| Length Scale Factors in Similar Shapes [Similarity and Congruency] [Geometry and Measure] | BIDMAS [Basic Arithmetic] [Number] | 1 | 1 | 1 | 1 |

Table 17: Full list of relationships found by *Random* in the Eedi topics dataset. Each row refers to one relationship (one edge). From left to right, the columns are the sending node (topic), the receiving node (topic), and the adjacency and orientation evaluations from each expert. For each topic, the brackets contain its parent level 2 and level 1 topics.

| | Adjacency | | | Orientation | | | Causal accuracy |
|---|---|---|---|---|---|---|---|
| | Recall | Precision | $F_1$-score | Recall | Precision | $F_1$-score | |
| PC | 0.312±0.072 | 0.508±0.077 | 0.381±0.057 | 0.123±0.042 | 0.202±0.069 | 0.136±0.063 | 0.211±0.065 |
| GES | 0.295±0.057 | 0.473±0.043 | 0.378±0.051 | 0.132±0.055 | 0.200±0.071 | 0.138±0.058 | 0.238±0.048 |
| NOTEARS (L) | 0.123±0.047 | 0.401±0.063 | 0.219±0.040 | 0.093±0.044 | 0.291±0.050 | 0.129±0.047 | 0.090±0.035 |
| NOTEARS (NL) | 0.222±0.048 | 0.434±0.052 | 0.293±0.050 | 0.157±0.068 | 0.332±0.061 | 0.228±0.044 | 0.189±0.041 |
| DAG-GNN | 0.332±0.071 | 0.413±0.059 | 0.354±0.072 | 0.262±0.049 | **0.336±0.060** | 0.285±0.051 | 0.257±0.066 |
| VISL | **0.698±0.052** | 0.588±0.042 | **0.635±0.049** | **0.417±0.065** | 0.314±0.069 | **0.359±0.059** | **0.615±0.074** |
| MMHC | 0.612±0.057 | **0.602±0.051** | 0.601±0.046 | 0.355±0.041 | 0.261±0.041 | 0.298±0.058 | 0.471±0.061 |
| Tabu | 0.332±0.042 | 0.461±0.053 | 0.390±0.047 | 0.121±0.052 | 0.198±0.054 | 0.128±0.051 | 0.240±0.050 |
| HillClimb | 0.291±0.054 | 0.452±0.060 | 0.361±0.051 | 0.134±0.049 | 0.196±0.061 | 0.130±0.050 | 0.221±0.044 |

Table 18: Structure learning results for the extended synthetic experiment described in Table 11. Each value is the mean and standard error over twenty-four datasets. In general, the results are qualitatively similar to those obtained in the synthetic experiment in the paper (recall Table 2), with VISL obtaining superior performance compared to the previous and the new baselines. Notice that the new baseline MMHC is close to VISL, being superior in adjacency-precision. However, as shown in Table 11, MMHC scales poorly.

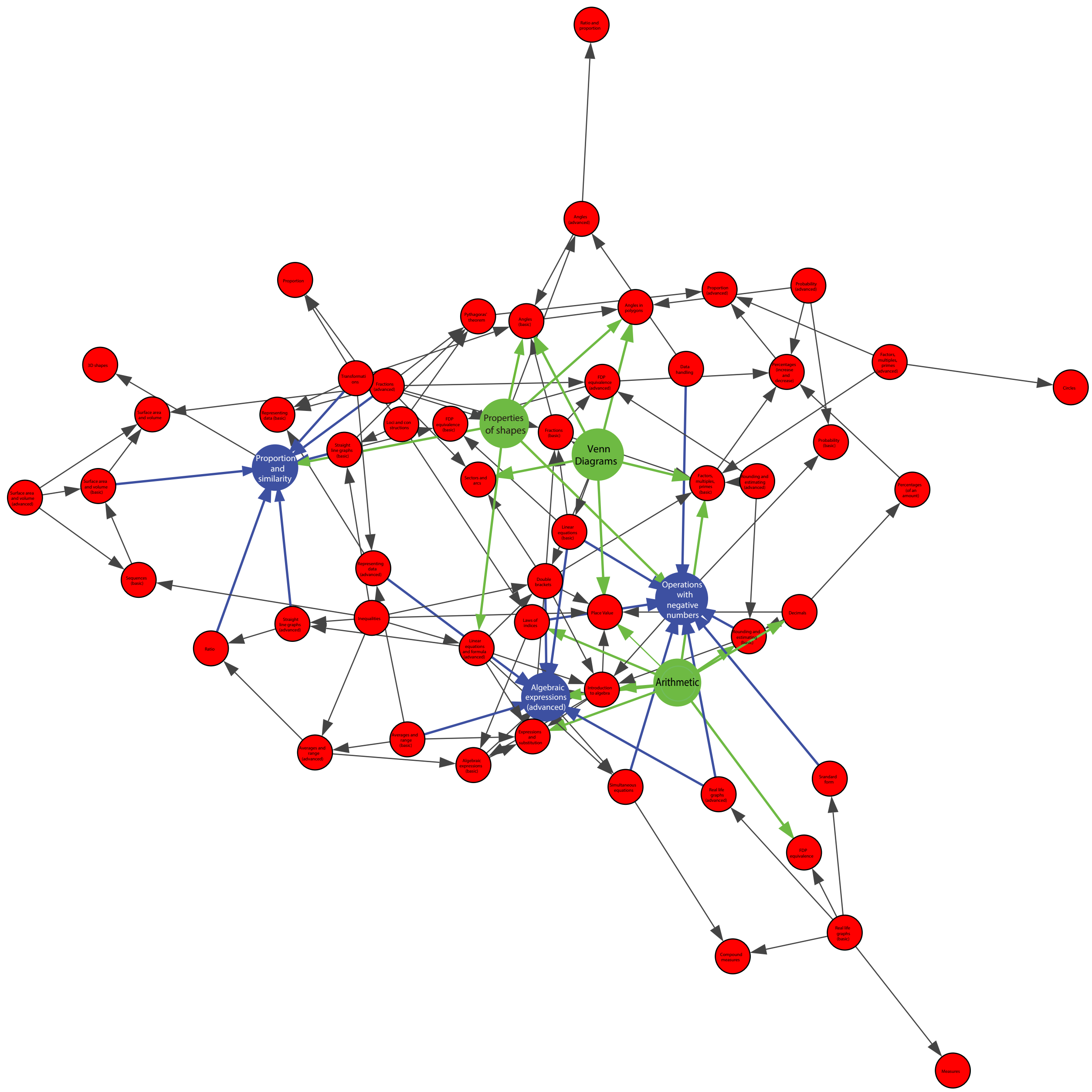


Figure 8: The structures discovered by VISL based on an alternative education dataset. In particular, the blue node represents the curriculum that should be taught later. On the other hand, the green node represents the fundamental topics that should be taught earlier.