

# Prerequisite Skill Structures in ASSISTments

Michel Desmarais      Sameer Bhatnagar

July 28, 2015

## Abstract

At the 2015 Artificial Intelligence in Education conference, Seth Adjei and Neil Heffernan presented their work on scrutinizing expert-defined prerequisite skill graphs. Using randomized controlled trials in PLACEments, the computer-adaptive-testing feature in the ASSISTments learning environment, they were able to identify some prerequisite skill arcs that were not supported by data.

This proposal outlines two alternative techniques that could be used to achieve the goal. We will start with a basic introduction to Partial Order Knowledge Structures (POKS), then try to mathematically formalize what a prerequisite skill might mean in this framework. The first proposed technique will be graph based, while the second will be built on interpreting probabilities during POKS structural learning.

## 1 POKS

Partial Order Knowledge Structures (POKS) determine which items in a test are prerequisite to others. POKS are derived from the theory of Knowledge Spaces, where such prerequisite relationships are written as  $A \rightarrow B$ , which means if a student got item A correct, they likely will get item B also correct; said differently, item B is a prerequisite of item A.

### 1.1 Working Example

Let us look at a classical dataset from Tatsuoka, which is well studied in de la Torre (*skill names not clear, as they are in de la Torre 2009, which is not accessible*).

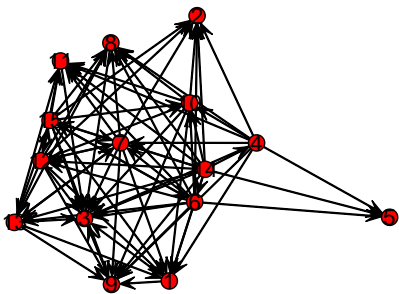
### 1.1.1 POKS Network Induction

##	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12	T13	T14
## 1	0	1	0	1	0	1	1	1	1	1	1	0	1	1
## 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
## 3	1	1	1	1	0	0	0	0	1	1	1	0	0	0
## 4	1	1	1	0	0	1	1	1	1	0	1	0	1	0
## 5	0	1	1	0	0	0	0	1	0	0	0	0	0	0
## 6	0	0	1	0	1	0	0	0	1	0	0	0	0	0
##	T15													
## 1	1													
## 2	1													
## 3	0													
## 4	1													
## 5	0													
## 6	0													
##	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12	T13	T14
## 531	0	0	0	0	0	0	1	0	0	0	0	0	0	0
## 532	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## 533	1	1	1	1	0	1	1	1	1	0	1	1	1	0
## 534	0	0	0	0	0	0	0	0	0	0	0	0	0	0
## 535	1	1	1	1	1	1	1	1	1	1	1	0	1	0
## 536	1	1	1	0	0	0	0	1	1	1	1	0	0	0
##	T15													
## 531	0													
## 532	0													
## 533	0													
## 534	0													
## 535	0													
## 536	0													

If we run the POKS code on this data, we get the following adjacency matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
2	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
4	1	1	1	0	1	0	1	1	1	1	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1
7	1	1	1	0	0	0	0	1	1	0	1	1	1	0	0
8	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	1	0	0	0	0	1	1	0	1	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	1	0	0	0	1	1	1	1	1	0	1	0	1
13	1	0	1	0	0	0	1	1	1	0	1	0	0	0	0
14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
15	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0

This structure can be visualized as follows:



What stands out is item three, which has no prerequisites (no edges leaving

node 5). Let us use this as a place to work out some numbers to demonstrate how the POKS structure is induced. We see that mastery of item 4 implies mastery of item 5 from the table and the graph. Let us check the contingency table for these two items.

```
##      item5
## item4   0   1
##      0 149 173
##      1  40 174
```

Now we can calculate  $P(item_5|item_4)$  which should be greater than some threshold, as should  $P(\neg item_5|\neg item_4)$ .

We must also ensure that the distributions over the two items are actually interacting with one another using a chi-square test.

This is repeated pairwise for all the items until a item-item structure is built, with directed edges indicating prerequisite relations.

### 1.1.2 Inference

POKS is already used at Polytechnique in an computer-aided adaptive test taken by new undergraduates to assess their math skills. The item-item structure is used as the basis for selecting the next question for any given student. However, everytime a student answers a question, their response provides evidence that can be used to update the probabilities of other items. This item-item

## 2 Prerequisite Skills

It is important to recognize that to go from item-item structures, to skill-skill structure, we need an item-skill mapping. This has been coined the Q-matrix, where each item may require one or more skills. Here is the expert-defined q-matrix from our dataset:

The five skills (columns) are

**QT1** performing basic fraction-subtraction operation

	QT1	QT2	QT3	QT4	QT5
T01	1	0	0	0	0
T02	1	1	1	1	0
T03	1	0	0	0	0
T04	1	1	1	1	1
T05	0	0	1	0	0
T06	1	1	1	1	0
T07	1	1	1	1	0
T08	1	1	0	0	0
T09	1	0	1	0	0
T10	1	0	1	1	1
T11	1	0	1	0	0
T12	1	0	1	1	0
T13	1	1	1	1	0
T14	1	1	1	1	1
T15	1	1	1	1	0

**QT2** simplifying/reducing

**QT3** separating whole number from fraction

**QT4** borrowing one from whole number to fraction

**QT5** converting whole

The two proposals below will try to derive the prerequisite graph of these skills. Each method takes a different approach to mathematically formalizing how a skill-prerequisite is modeled.

## 2.1 Deriving Prerequisite skills - proposal 1

Multiplication of the student-response matrix, with the Q-matrix, gives what is called the skill-mastery matrix, wherein we can see which students have mastered which skills.

##		QT1	QT2	QT3	QT4	QT5
##	[1,]	11	8	10	8	3
##	[2,]	14	8	12	9	3
##	[3,]	7	2	5	3	2
##	[4,]	10	6	7	5	0
##	[5,]	3	2	1	1	0

```
## [6,] 2 0 2 0 0
##      QT1 QT2 QT3 QT4 QT5
## [531,] 1 1 1 1 0
## [532,] 0 0 0 0 0
## [533,] 11 6 8 6 1
## [534,] 0 0 0 0 0
## [535,] 11 6 9 6 2
## [536,] 7 2 4 2 1
```

This is a projection of the student test data onto the skills: the higher the value for a student, the greater our confidence that they mastered that skill. This can be normalized by the maximum possible score the student could have had for any particular skill, which gives a skill-mastery-probability matrix:

```
##      QT1 QT2 QT3 QT4 QT5
## [1,] 0.79 1.00 0.83 0.89 1.00
## [2,] 1.00 1.00 1.00 1.00 1.00
## [3,] 0.50 0.25 0.42 0.33 0.67
## [4,] 0.71 0.75 0.58 0.56 0.00
## [5,] 0.21 0.25 0.08 0.11 0.00
## [6,] 0.14 0.00 0.17 0.00 0.00
##      QT1 QT2 QT3 QT4 QT5
## [531,] 0.07 0.12 0.08 0.11 0.00
## [532,] 0.00 0.00 0.00 0.00 0.00
## [533,] 0.79 0.75 0.67 0.67 0.33
## [534,] 0.00 0.00 0.00 0.00 0.00
## [535,] 0.79 0.75 0.75 0.67 0.67
## [536,] 0.50 0.25 0.33 0.22 0.33
```

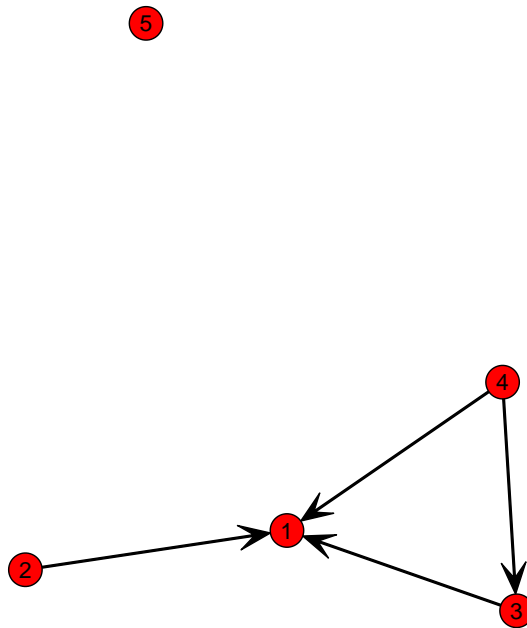
Now if we consider each skill as an item, and round the probabilities from the previous matrix, we get something that looks a lot like another item-response matrix (just like the raw test data we started with).

Every probability that was above 0.5 is rounded to 1 (the student mastered that skill). Otherwise, the student gets a 0 for that skill.

We can now run the same POKS code as before, but on this new skill mastery matrix: the difference is that now instead of determining which *items* are prerequisites of each other, we will get relationships describing mastery of which **skills** imply other **skills**.

##		[,1]	[,2]	[,3]	[,4]	[,5]
##	[1,]	0	0	0	0	0
##	[2,]	1	0	0	0	0
##	[3,]	1	0	0	0	0
##	[4,]	1	0	1	0	0
##	[5,]	0	0	0	0	0

Once again, this can be visualized as follows (remembering that each node now represents a skill)



The threshold probability was raised to 0.99 for this simulation. Comparing this to the original list of skills, we believe there is sufficient face-value to this approach to try further validation.

## 2.2 Deriving Prerequisite skills - proposal 2

Given latent skill  $S_{k_A}$ , associated with items  $\Omega_A = \{X_{1S_{k_A}}, X_{2S_{k_A}} \dots X_{nS_{k_A}}\}$ , we can find the sum of the KL-divergence's for each pair of item distributions in order to measure how far apart these items are from each other.

$$ItemDistance_{S_{k_A}} = \sum_i^n \sum_j^n D_{KL}(X_{iS_{k_A}} \parallel X_{jS_{k_A}}), i \neq j$$

Assume now that there are other latent skills,  $S_{k_B}, S_{k_C}, S_{k_D}$ , etc., each which is associated with its own set of items,  $\Omega_B, \Omega_C, \Omega_D$ , etc. Some of these skills should be mastered before  $S_{k_A}$ . The set of prerequisite skills can be defined as those which minimize  $ItemDistance_{S_{k_A}}$ , given mastery of the items related to them.

$$\Omega^* = \arg \min_{\Omega} ItemDistance_{S_{k_A}}$$

### 2.2.1 proposal 2-modified

Instead of KL Divergence, we could also sum the entropy  $H_{X_{jS_{k_A}}}$  of each the items related to skill A, which be

$$H_{TS_{k_A}} = \sum_{j=1}^n H_{X_{jS_{k_A}}}$$

$$\Omega^* = \arg \min_{\Omega} H_{TS_{k_A}}$$

## 3 Validation

Once a prerequisite skill structure is determined, it can be validated by seeing if running a simulation of a computer adaptive test, knowledge of the skill structure would improve predictions of performance over POKS alone.