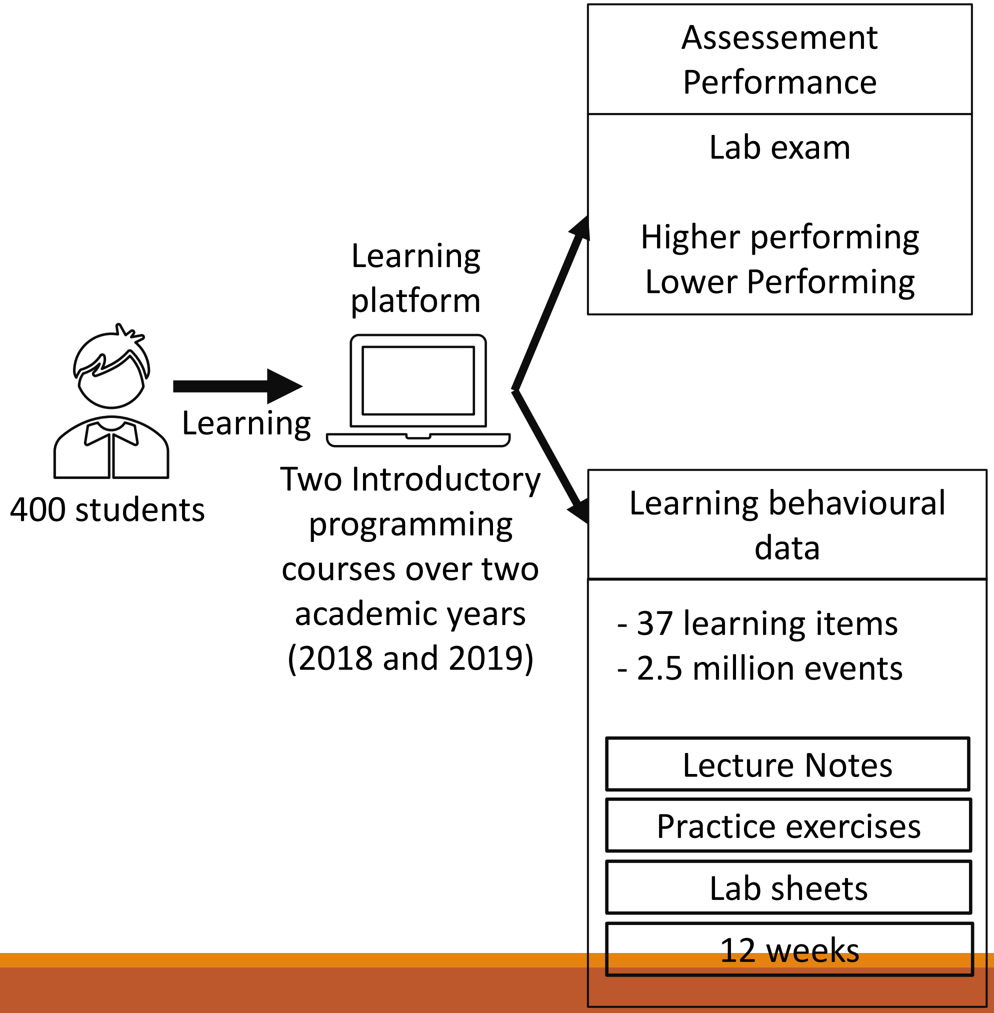
Community analysis and learning outcome prediction based on learning behaviours in the context of programming education

TAI MAI, MARIJA BEZBRADICA, MARTIN CRANE,
SCHOOL OF COMPUTING, DUBLIN CITY UNIVERSITY

Introduction

- The failure rate in introductory programming modules has been reported to be 28% on average, with a huge variation from 0% to 91% (Bennedsen & Caspersen, Michael, 2019)
- Learning behaviours tend to be correlated with students' performance in programming education (Carter & Hundhausen, 2017)
- The learning behaviours in using material items, however, has not been commonly investigated (Li & Tsai, 2017)
- Problem with noise in the dataset.

Context of the study



Example of event log of student $s1$ on two days in week 5.

Trace id	Event Item	Timestamps	Student id
1	Labsheet 5	2018-08-12 14:30:00	s1
1	Labsheet 5	2018-08-12 14:35:00	s1
1	Lecture 5	2018-08-12 14:36:00	s1
1	Labsheet 5	2018-08-12 14:45:00	s1
1	Lecture 5	2018-08-12 14:49:00	s1
1	Labsheet 5	2018-08-12 14:50:00	s1
1	Labsheet 5	2018-08-12 15:00:00	s1
1	...	...	s1
2	Labsheet 5	2018-08-13 11:54:00	s1
2	Practice 5	2018-08-13 11:59:00	s1
2	...	...	s1

Datasets information.

Dataset	Number of students	Number of events	Average events per student
Course#1-2018	126	1,054,394	8368
Course#1-2019	164	1,484,297	9050
Course#2-2018	62	211,855	3417
Course#2-2019	52	200,006	3846

Behavioural features

Example of event log of student $s1$ on two days in week 5.

Trace id	Event Item	Timestamps	Student id
1	Labsheet 5	2018-08-12 14:30:00	s1
1	Labsheet 5	2018-08-12 14:35:00	s1
1	Lecture 5	2018-08-12 14:36:00	s1
1	Labsheet 5	2018-08-12 14:45:00	s1
1	Lecture 5	2018-08-12 14:49:00	s1
1	Labsheet 5	2018-08-12 14:50:00	s1
1	Labsheet 5	2018-08-12 15:00:00	s1
1	...	...	s1
2	Labsheet 5	2018-08-13 11:54:00	s1
2	Practice 5	2018-08-13 11:59:00	s1
2	...	...	s1

Example of student-event item data matrix

StudentId	Lecture1	Labsheet1	Practice2	...
s1	5	7	6	...
s2	24	14	34	...
s3	12	54	0	...
...	...	...	...	...

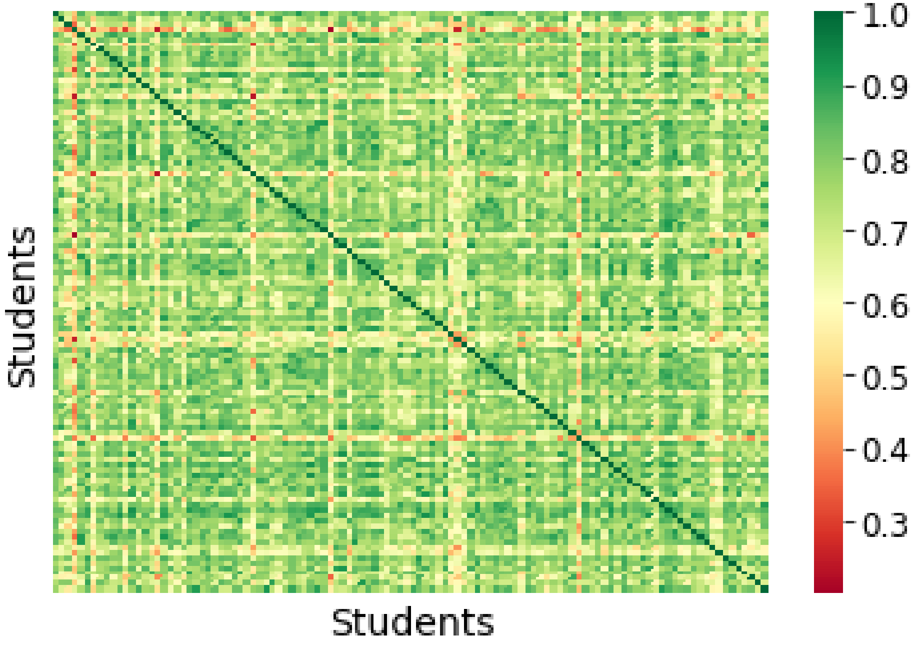
Example of transition-student data matrix

Transition	s1	s2	s3	s4	...
Lecture1-Lecture1	4	5	10	23	...
Lecture1-Labsheet1	0	14	9	12	...
Labsheet1-Practice1	12	6	0	21	...
Labsheet1-Lecture1	16	25	0	5	...
...	...	...	...	...	...

Noise and Trend effect

Example of transition-student data matrix

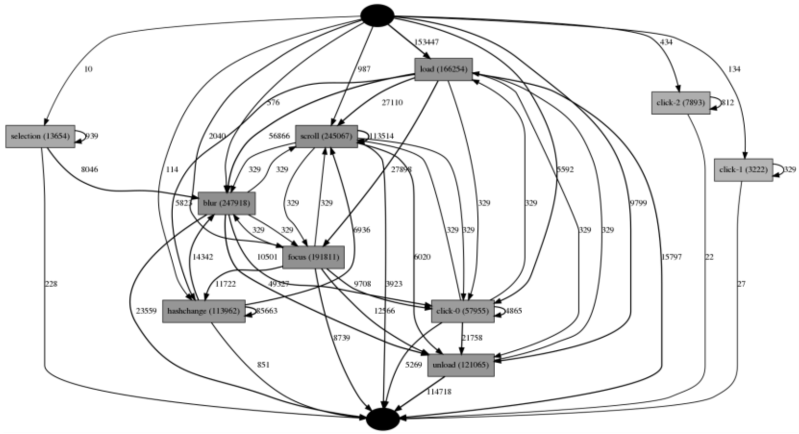
Transition	s1	s2	s3	s4	...
Lecture1-Lecture1	4	5	10	23	...
Lecture1-Labsheet1	0	14	9	12	...
Labsheet1-Practice1	12	6	0	21	...
Labsheet1-Lecture1	16	25	0	5	...
...	...	...	...	...	...



Correlation matrix

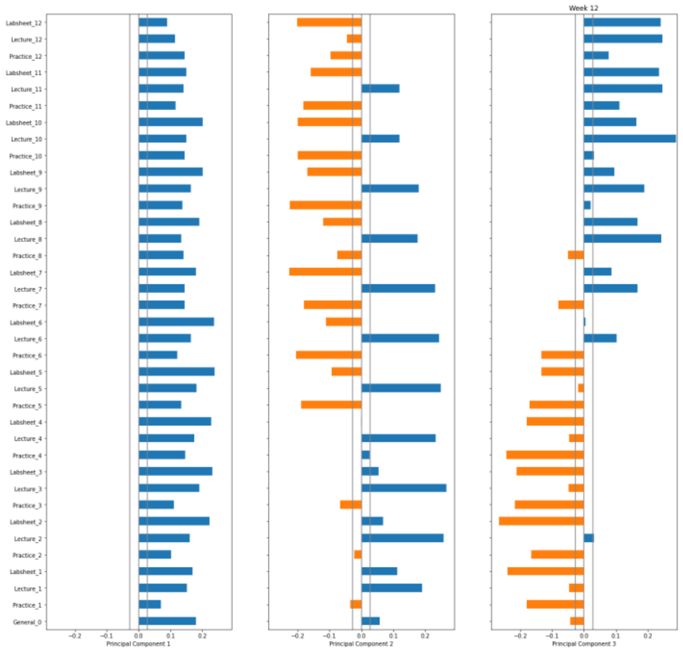
- Students freely interactive with the learning platform

- High correlation between students learning behaviours



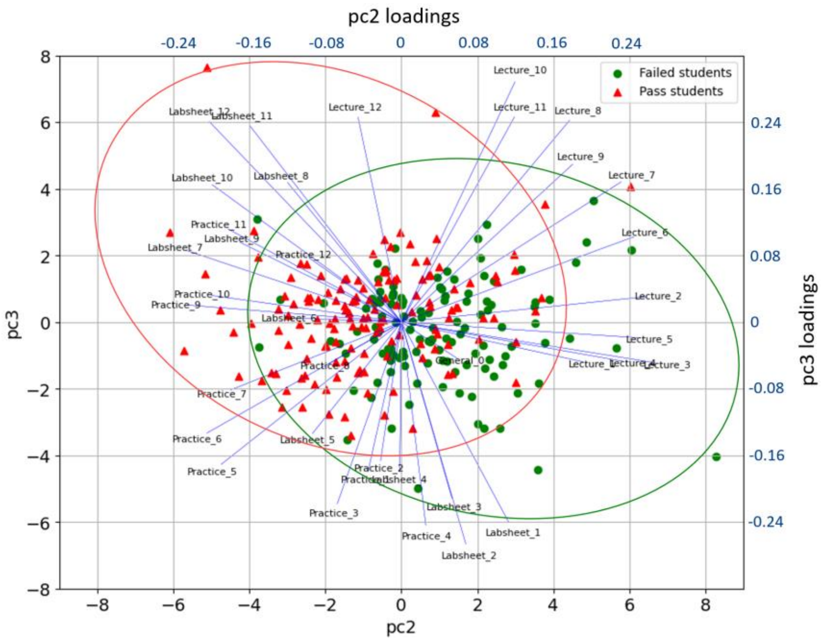
Process mining approach

Noise and Trend effect – Investigation with PCA



PC1 PC2 PC3

Eigenvectors: PC1 loadings are all positive => Possible Noise & Trend effect



Biplot of PC2 & PC3 reveals differences of learning behaviours between higher-performing and lower-performing cohorts

Research objective

- Community Analysis
- Learning outcome prediction
- Deal with the problem of noise and trend in the dataset

Random Matrix Theory

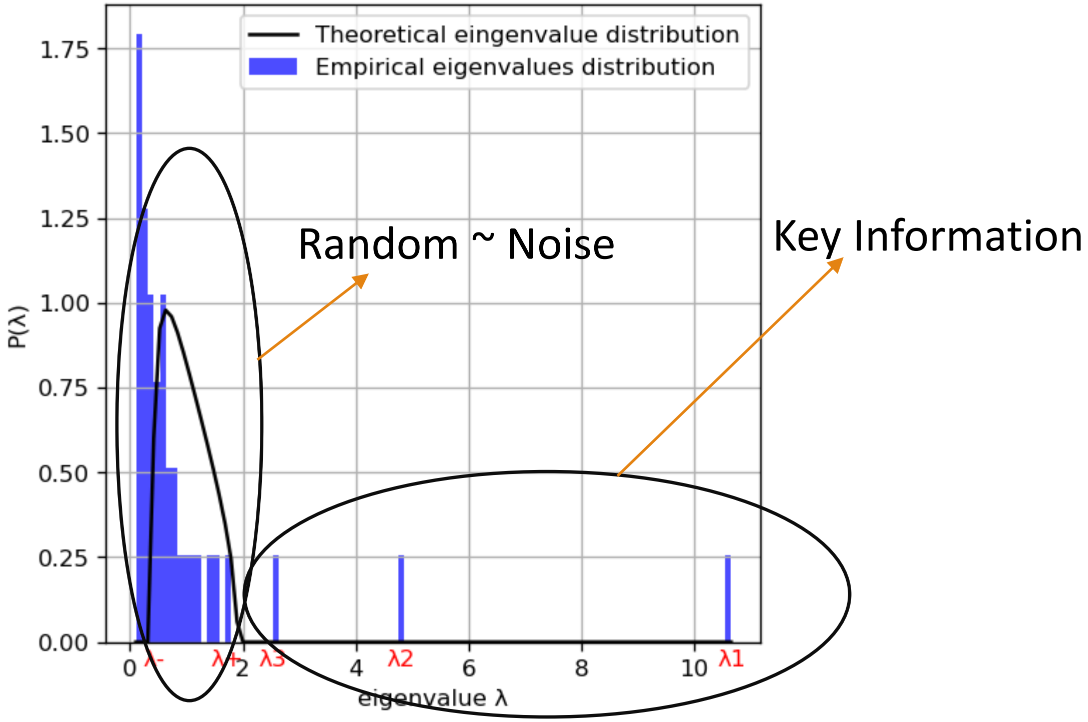
Given a random matrix *mxn* **A** such that $Q = m/n > 1$ is fixed. **R** is a correlation matrix of **A**

Marcenko-Pastur probability density function of eigenvalue λ of **R** is given by:

$$P_R(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

where $\lambda_- \leq \lambda \leq \lambda_+$, λ_- and λ_+ are lower and upper bounds, (i.e. minimum and maximum), eigenvalues of R respectively, given by:

$$\lambda_\pm = \sigma^2\left(1 \pm \sqrt{\frac{1}{Q}}\right)^2$$



=> Clean correlation matrix

=> Clean dataset

Clean correlation matrix

Eigenvalue clipping

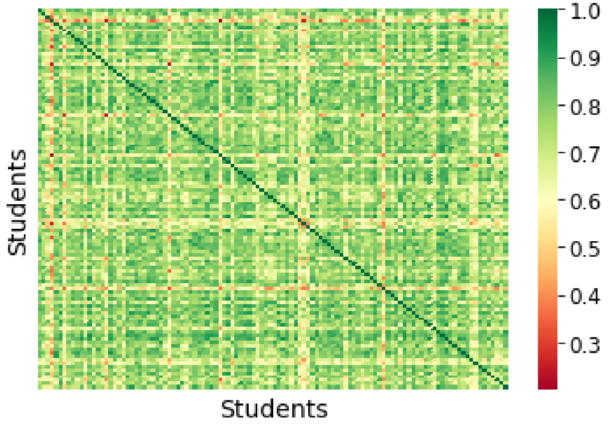
Let $\lambda_1, ..., \lambda_N$ be the set of all eigenvalues of **C** and $\lambda_1 > ... > \lambda_N$, and i be the position of the eigenvalue such that $\lambda_i > \lambda_+ and \lambda_{i+1} \leqslant \lambda_+$.

Then we set

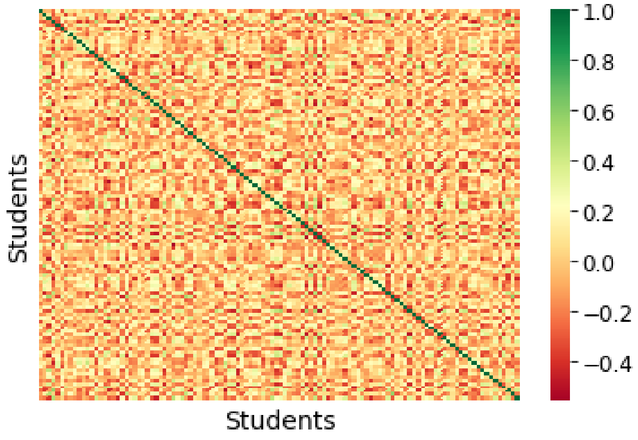
$$\lambda_j = 1/(N-i) \sum_{k=i+1}^{N} \lambda_k, \quad (7)$$

→ $C_{denoised}$ → $C_{cleaned} = C_{denoised} - W_1 V_1 W_1^T$

- Where **C** is the correlation matrix of the standardised dataset **G**
- ***W1*** and ***V1*** are the first eigenvector and eigenvalue of **C**



cleaning →



Clean dataset

Let $\lambda_1, ..., \lambda_N$ be the set of all eigenvalues of **C** and $\lambda_1 \geq ... \geq \lambda_N$, and k be the position of the eigenvalue such that $\lambda_k > \lambda_+$ and $\lambda_{k+1} < \lambda_+$. We note that λ_1 refers to the largest eigenvalues and the first principal component. The clean dataset $\widehat{G}$ can be constructed as follows:

$$\widehat{G} = \sum_{i=1}^{n} v_i x_i - \alpha \sum_{i=k}^{n} v_i x_i - \beta v_1 x_1 \qquad (9)$$

v_i :eigenvector *i-th* (loadings of component i-th)
x_i :scores of component *i-th*

$\alpha \in [0, 1]$
$\beta \in [0, 1]$
Configurable parameters:
how much noise and trend we want to clean

Fully re-constructed dataset **G** from component scores and loadngs

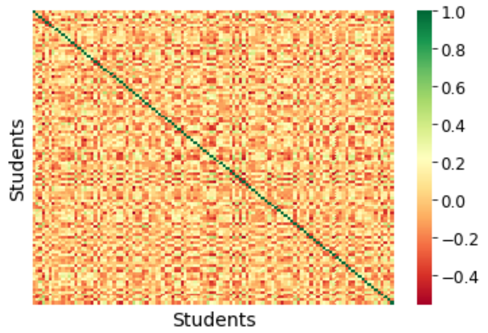
Noise part

Trend part

Community Analysis

- Construct a graph from distance matrix
- Girvan-Newman community detection algorithm
- Labelling each community based on the number of higher/lower-performing students in the community
- Comparison analysis

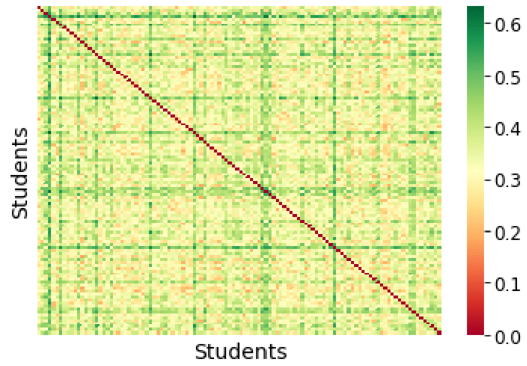
Community Analysis



Cleaned correlation matrix **C**

Distance of learning behaviours between student *i* and *j*:

$$\mathbf{D}_{ij} = \sqrt{0.5 * (1 - \mathbf{C}_{ij})}$$



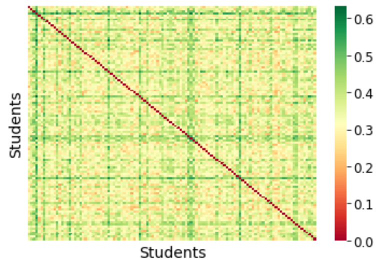
Distance matrix **D**

$D_{ij} \in [0, 1]$

$D_{ij} \rightarrow 1$ Learning behaviours between two students are more different

$D_{ij} \rightarrow 0$ Learning behaviours between two students are more similar

Graph construction

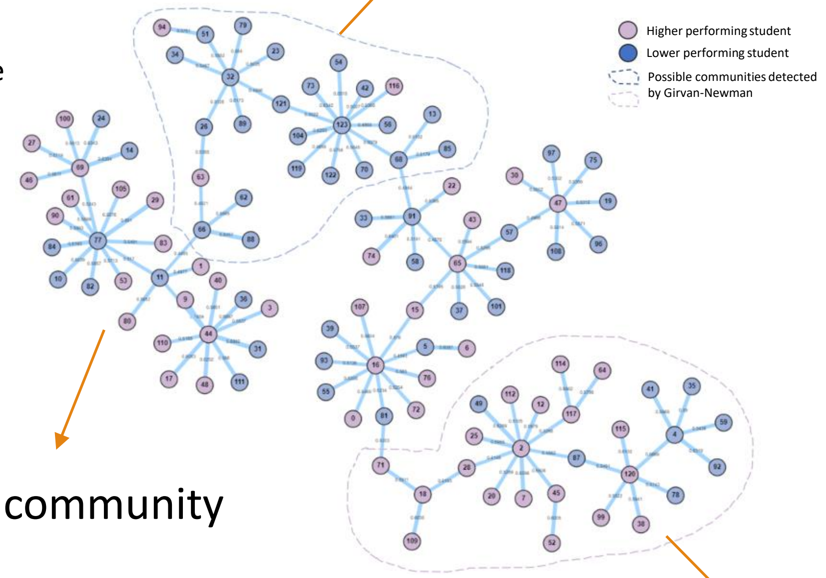


Distance matrix **D**

Minimum spanning tree

Girvan-Newman community detection

Lower performing community (>70% of lower-performing students)



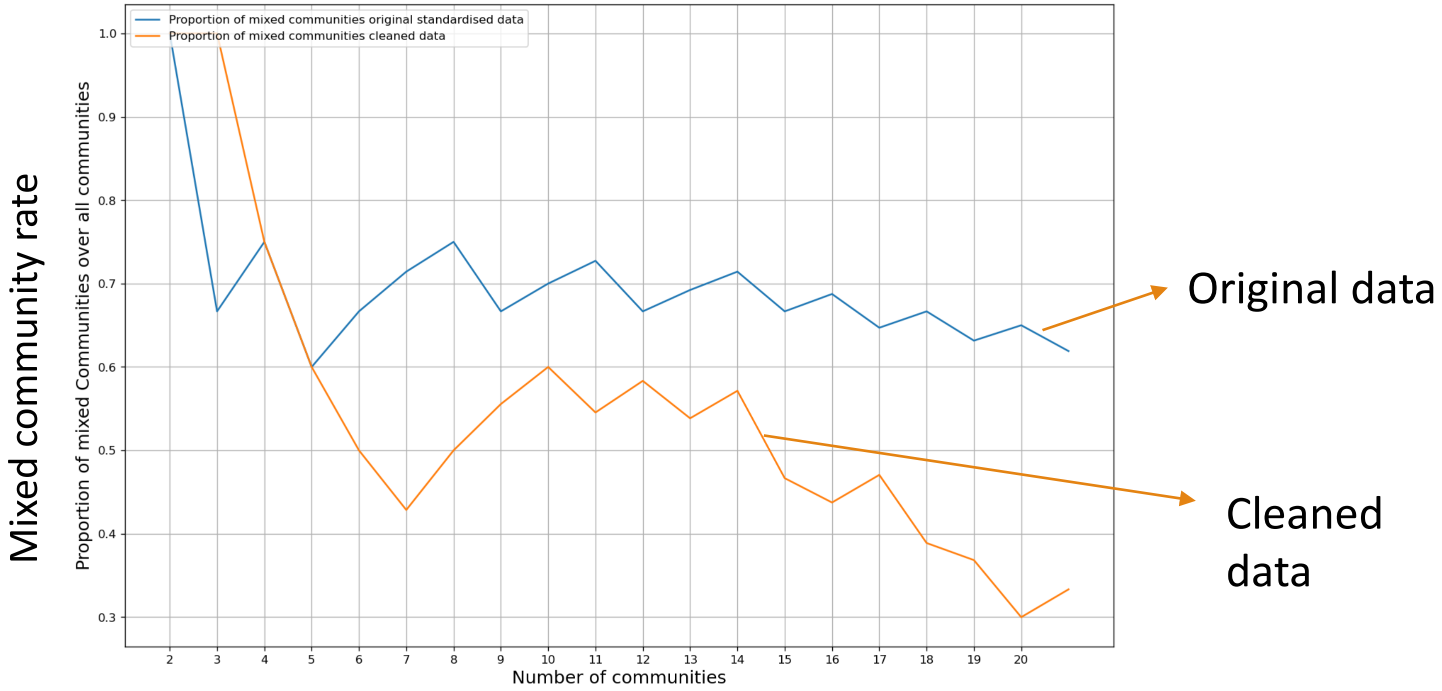
Mixed community

Higher performing community (>70% of higher-performing students)

$$\textit{mixed community rate} = \frac{\textit{number of mixed communities}}{\textit{total number of detected communities}}$$

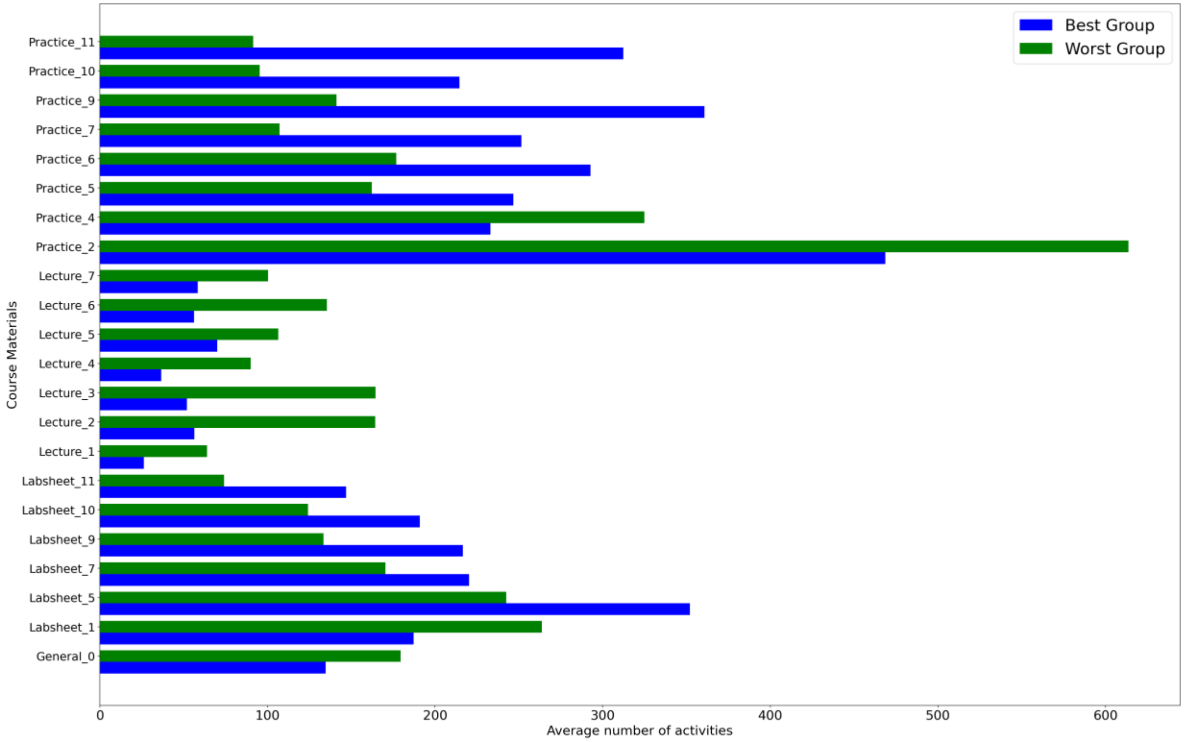
We want less mixed communities and more lower/higher-performing communities

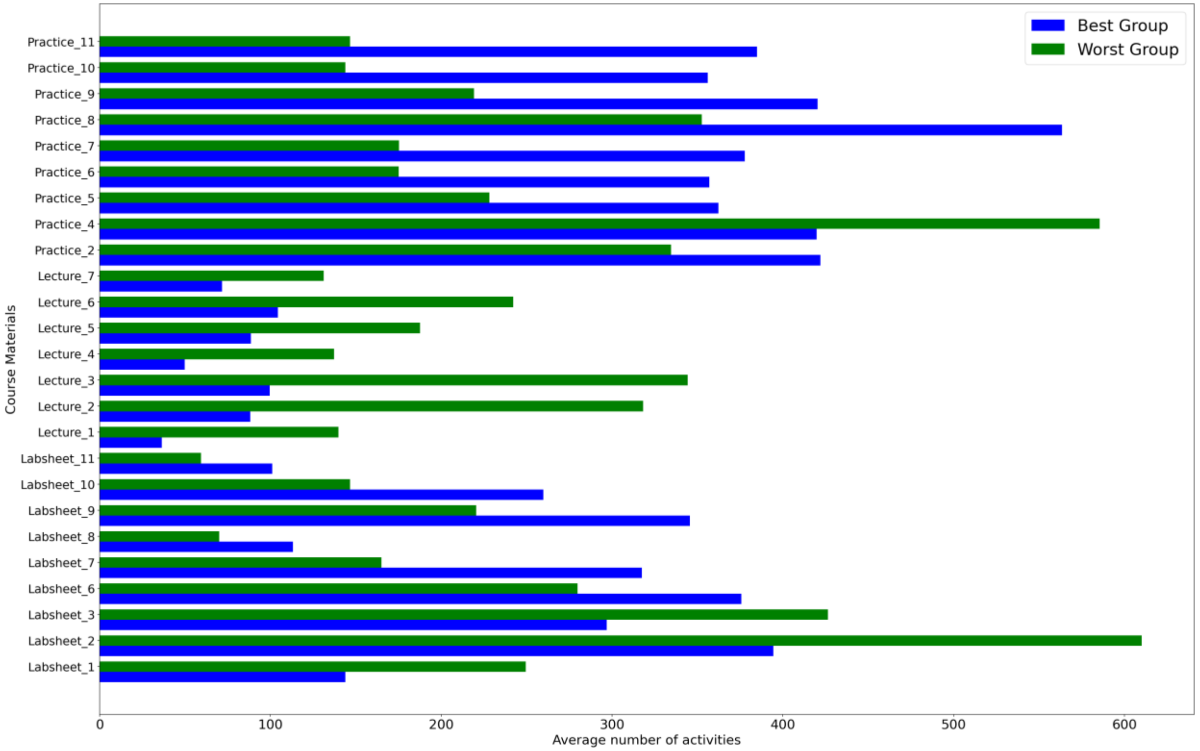
Community Analysis



Comparison between original data and cleaned data when apply Girvan-Newman algorithm

Community Analysis
Compare best and worst performing communities

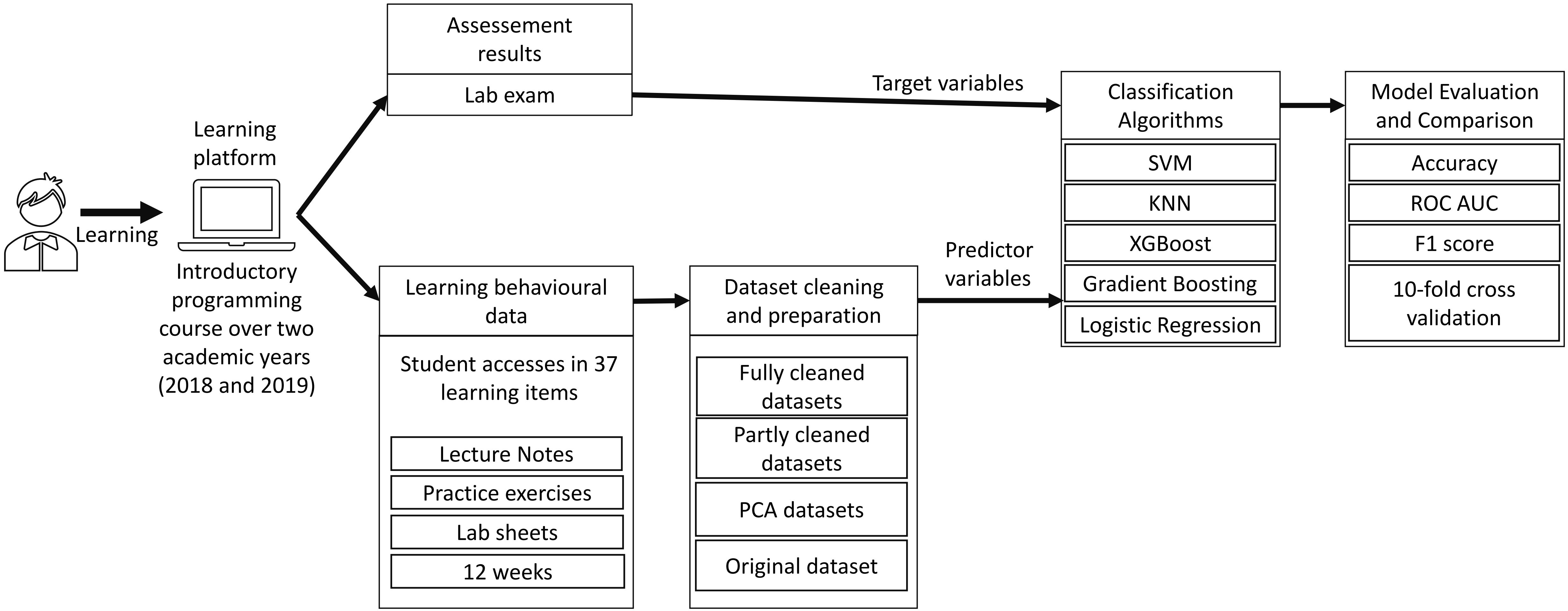




Lower performing group: doing less practice and lab instruction, reading more lecture notes

Higher performing group: doing more practice and lab instructions, reading less lecture notes

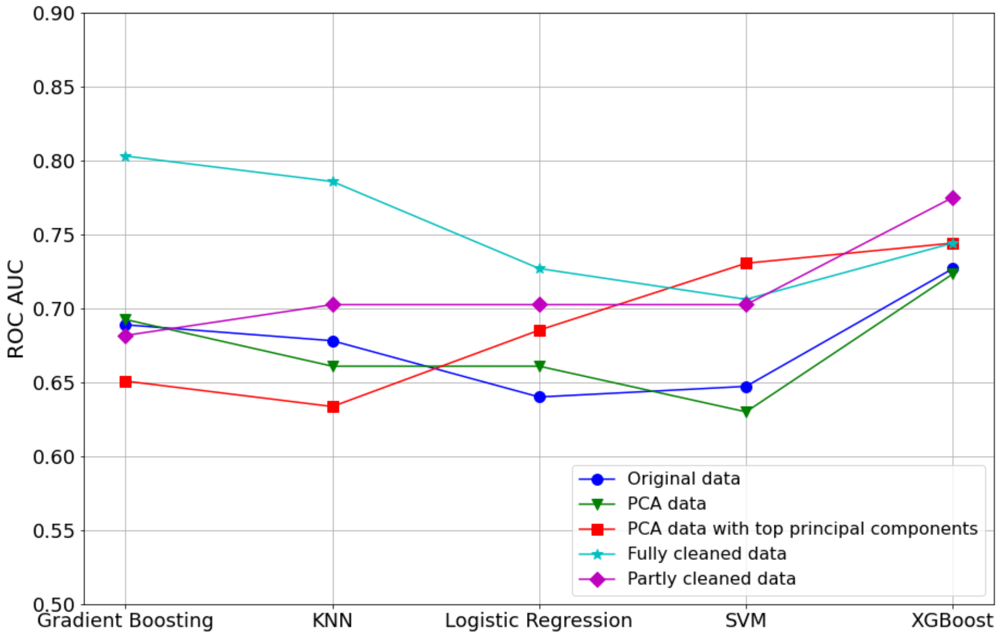
Learning outcome prediction

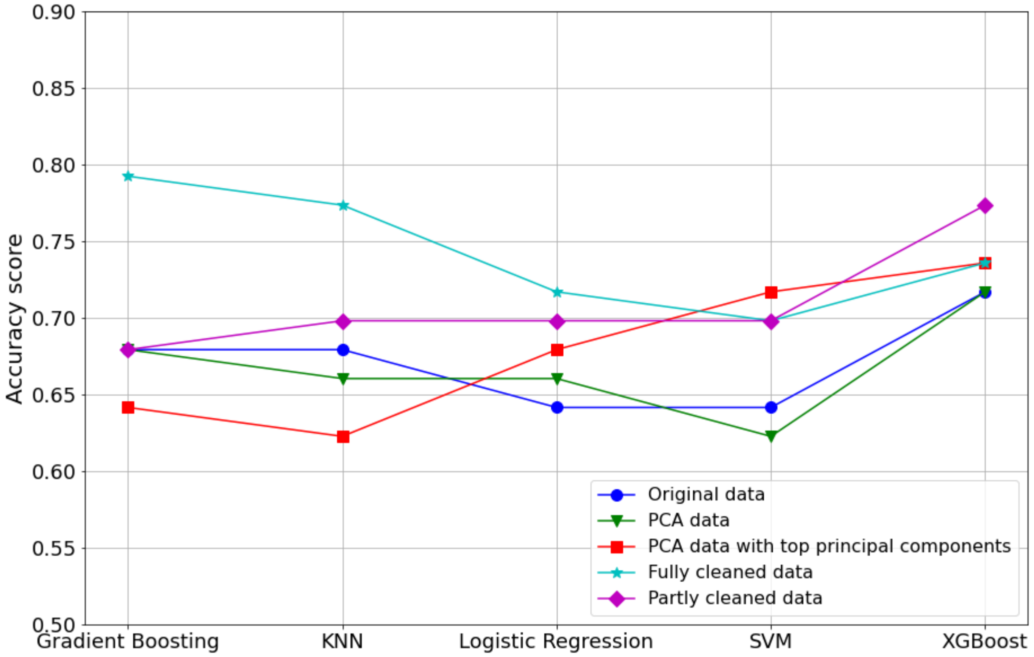


Example of student-event item data matrix

StudentId	Lecture1	Labsheet1	Practice2	...
s1	5	7	6	...
s2	24	14	34	...
s3	12	54	0	...
...	...	...	...	...

Prediction result





Summary

- **Findings**
  - Extracted learning behavioural data
  - Utilised Random Matrix Theory (RMT) in Educational context to separate the key information from the noise in the dataset
  - Community Analysis and Learning performance prediction
  - Cleaned data can help to cluster more informed communities and improve prediction models
- **Future work**
  - More data features included (e.g. time and sequences of learning activities)
  - Compare different ways of constructing graph
  - Test with more datasets

THANK YOU FOR YOUR ATTENTION!