Neo4j4Us

Graph Databases & Bayesian Networks to assist students with planning their curriculum

Aims

Prediction Component

1. Create a generic system for Bayesian Network (BN) manipulation in Neo4j
2. Create a BN system for predicting probability of a student passing
3. Calculate model accuracy
4. Cluster students by aptitude and work ethic

Advisory Component

Compare Neo4j to MySQL for:

1. Checking degree constraints
2. Effectively identifying similar students

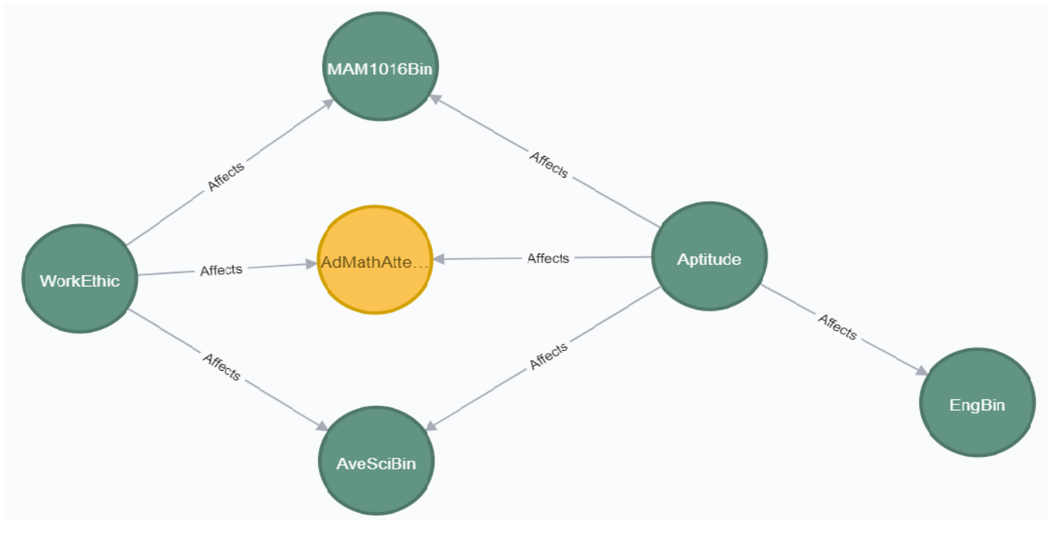
Process

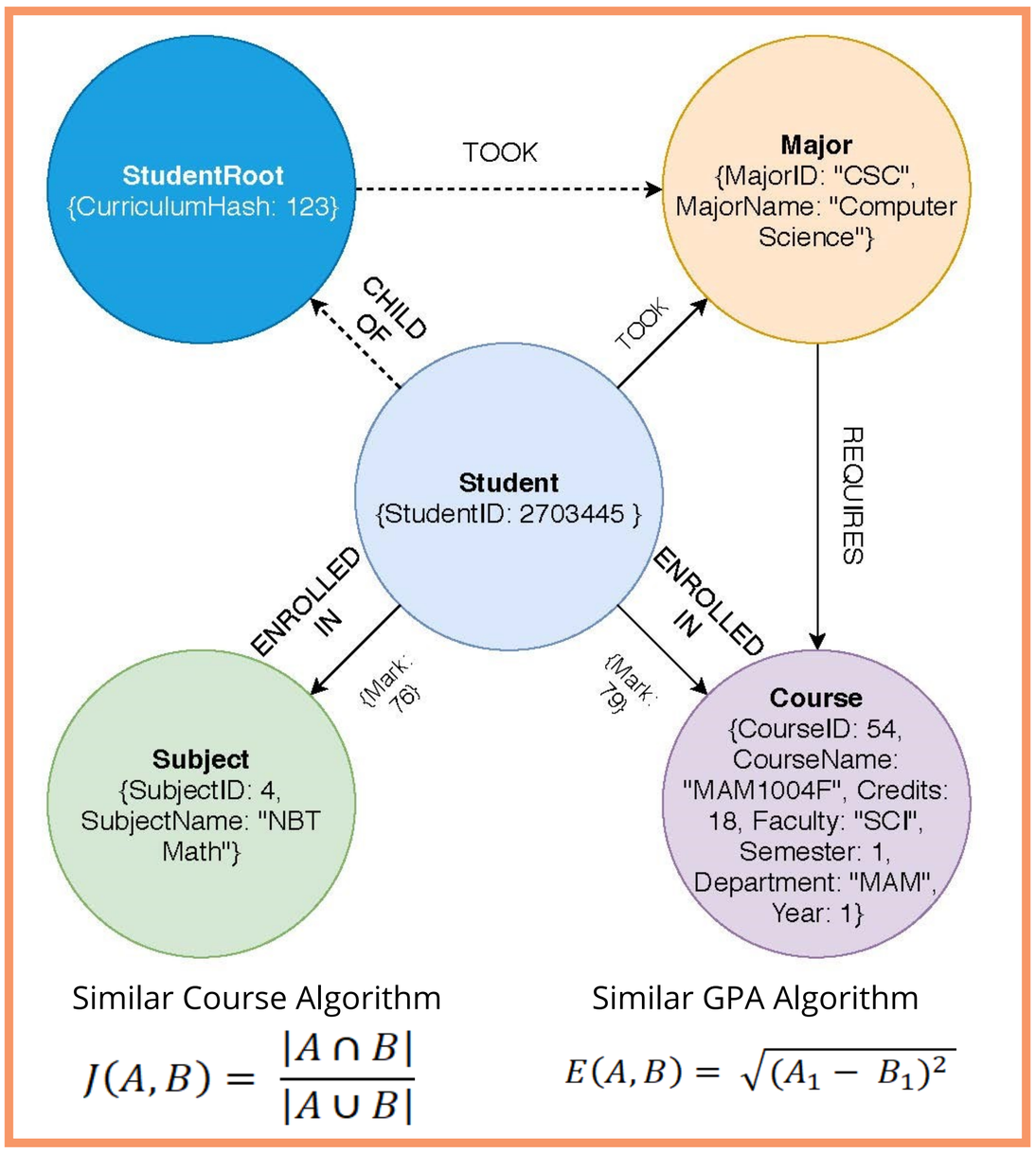
1. Upload AtRisk.DNE BN file to be parsed and transformed into a Neo4j graph
2. Identify students at risk of not completing their degree within 3 years
3. Perform testing using a hold-out set of student data
4. Use student clusters to create ‘teams’ of students likely to increase academic performance

1. Check degree constraints in graph
2. Find similar students according to enrolled courses and GPA
3. Compare the performance of Neo4j and MySQL in small & large DB’s

Design

```
+-------------------------------------------------------------------------------+
| Instructions: Neo4j4us can(currently) perform three tasks.                    |
+-------------------------------------------------------------------------------+
| (1) You can upload a raw Bayesian Network file in the form of a .dne file by  |
| entering 1 when asked "What would you like to do" and providing the complete  |
| file path to that file so Providentia can update its database.                |
+-------------------------------------------------------------------------------+
| (2) You can access an existing Bayesian Network in the database and perform   |
| inference by entering 2 when asked "What would you like to do" and then       |
| following the given instructions.                                             |
+-------------------------------------------------------------------------------+
| (3) You can view all Bayesian Networks in our library by entering 3 when      |
| asked "What would you like to do"                                             |
+-------------------------------------------------------------------------------+
```





Similar Course Algorithm

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Similar GPA Algorithm

$$E(A,B) = \sqrt{(A_1 - B_1)^2}$$

Results

1. Clusters used to group students can be used to improve teamwork performance
2. Testing with student-data holdout set yielded a BN accuracy of 89.41%
3. More experimentation needed to assess full potential of Neo4j used in Machine Learning
4. Any .DNE file can be parsed and encoded in a Neo4j database to be used in a local environment for prediction using an inference engine

1. Simpler and more efficient to implement constraint-checking in if-statements instead of graph queries
2. MySQL queries outperformed Neo4j in small databases. MySQL query performance degraded in larger databases—JOINing massive tables is too expensive. Neo4j performance was unaffected by DB size.

DB Size	Avg Query Execution Speed	
	MySQL	Neo4j
3000 Students	0.014 s	0.031 s
1M Students	7.224 s	0.035 s

Project Members:
Josh Redelinghuys (RDLJOS002@myuct.ac.za)
Edwin Kassier (KSSRUB001@myuct.ac.za)

Supervised By:
Sonia Berman (sonia@cs.uct.ac.za)
Deshen Moodley (deshen@cs.uct.ac.za)

University of Cape Town
Computer Science Dept.
dept@cs.uct.ac.za