Influence of Educational Resources on Employment Outcomes across Different Socioeconomic Backgrounds

Group 10 [Xiaohan Wang, Laurynas Jagutis, Guangyu Li]

Research Question

Explore the influence of *educational* resources on *employment* outcomes across different socioeconomic backgrounds.

Client context The client is the *government* that wants to research whether *investing in educational resources* will *improve economic mobility* in the country, by comparing progress in other countries

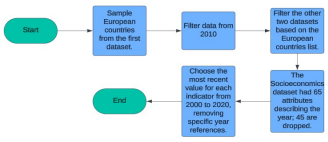
Data Sources

1. **Country Socioeconomic Status Scores**
   The Kaggle dataset on Country Socioeconomic Status Scores, Part II – the dataset contains estimates of the socioeconomic status (SES) position of each of 149 countries covering the period 1880-2010.
2. **The World Bank Dataset on Unemployment**
   Total (% of total labor force) (modeled ILO estimate) - it holds the unemployment % of total labor force of every country for 2023.
3. **The World Bank Dataset on Education Statistics**
   Describes itself as holding over 4,000 internationally comparable indicators that cover education access, progression, completion, literacy, teachers, population, and expenditures. However, this dataset is very *scarce* and contains thousands of empty values.

Data Exchange

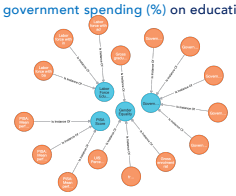
Data Cleaning

- **Sampled** European countries from the first dataset.
- **Filtered** data from 2010
- **Filtered** other two datasets based on the European countries list.
- The Socioeconomics dataset had 65 attributes describing the year; 45 were **dropped**.
- Chose the **most recent value** for each indicator from 2000 to 2020, removing specific year references.



Indicator Selection

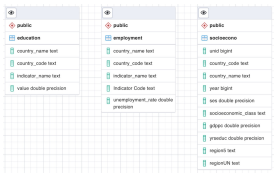
- The choice of relevant indicators was the main task for the third dataset. It consisted of 4000 of them and we had to narrow it down to a few.
  1. Remove all the indicators where at least one of our 18 European countries was missing a value.
  2. Investigated and chose the indicators which best represented the educational resources of a country, such as, government spending (%) on education.



Educational Resource Indicator Categories	Explanation	Member Indictors
Government Expenditure	The number of financial resources that a government allocates towards the education sector.	Government Expenditure as %, etc
Gender Equality	Measures the extent to which males and females have equal access to educational opportunities and resources.	Enrolment in primary education, both sexes (number), Enrolment in primary education, female (number).
Labor Force Education	This indicator focuses on the education levels within the labor force. It tracks the percentage of workers with various educational qualifications.	Labor force with advanced education (% of total)
PISA Score	The PISA (Programme for International Student Assessment) score assesses the educational performance of 15-year-olds in reading, mathematics, and science.	PISA: 15-year-olds by science proficiency level (%). Level 1A

Knowledge Graph Design

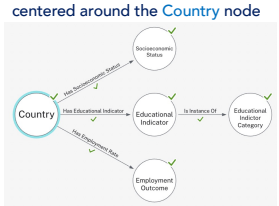
Relational Data Schema



A peek into our relational dataset

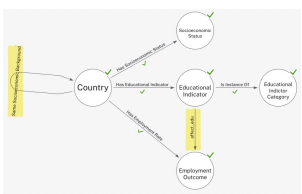
- Primary Keys
  - *country, [country, indicator]*
- Lack of intertwined foreign key
- Lack of topological structure
  - Could be merged into one not-so-large table

Initial Proposal:
centered around the Country node



Knowledge Graph Schema

- Derive Implicit Knowledge between nodes



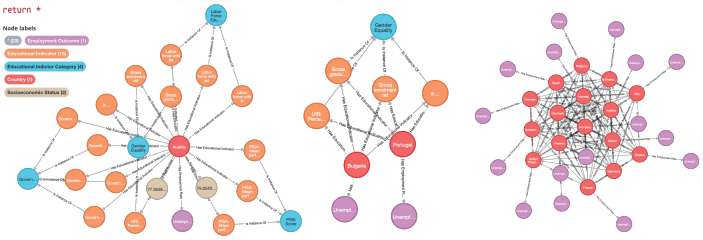
- Built with neo4j graph database
- and Cypher Query Language

- Entity Identification:
  - Countries, Indicators, Categories of Indicator, Employment Outcome, Socioeconomic Status
- Property Extraction
  - Add values of indicators as the properties of the nodes

Graph Queries and Data Analysis

Explanatory Graph Query: with our KG, various kinds of queries could be made

```
// Query on a single country
MATCH (n:Country {country_name: 'Austria'}) -[r]-> (i:`Educational Indicator`) -[r2]-> (ic:`Educational Indictor Category`),
(n:Country {country_name: 'Austria'}) -[r3]-> (e:`Employment Outcome`),
(n:Country {country_name: 'Austria'}) -[r4]-> (s:`Socioeconomic Status`)
return *
```

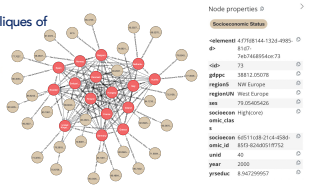



```
// Query Countries that connect to a Socioeconomic Status which has property "socioeconomic_class="Middle(semi-per)",
// their employment outcomes and educational indicators that is an istance of "Gender Equality"
MATCH (n:Country) -[r]-> (s:`Socioeconomic Status`),(n:Country) -[r2]-> (e:`Employment Outcome`),
(n:Country) -[r3]-> (i:`Educational Indicator`) -[r4]-> (ic:`Educational Indicator Category` {category: "category: Gender Equality"})
WHERE s.socioeconomic_class = 'High(core)' RETURN n, e, i, ic
```

Derive New Relations: Implicit knowledge mining

Besides what is presented in the data source, it is possible to form cliques of countries that have similar socioeconomic status in the KG

```
// Creates a new relation between countries that have the same socioeconomic status: High(core)
MATCH (n:Country) -[r]-> (s:`Socioeconomic Status`)
WHERE s.socioeconomic_class = 'High(core)' WITH n
MATCH (n2:Country) -[r2]-> (s2:`Socioeconomic Status`)
WHERE s2.socioeconomic_class = 'High(core)' AND n <> n2
MERGE (n)-[r3:Similar_Socioeconomic {socioeconomic_class: 'high(core)'}]-(n2)
```



Data Analysis

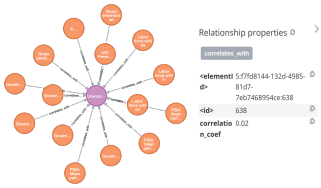
What kinds of educational indicators accounts most for the employment outcome?

	category	indicator_name	correlation_coef	indicator_code
0	Government Expenditure	Government expenditure on education as % of GD...	-0.24	SE.XPD.TOTL.GD.ZS
1	Government Expenditure	Government expenditure on secondary education ...	-0.40	UIS.XGDP.23.FSGOV
2	Government Expenditure	Government expenditure on tertiary education a...	-0.12	UIS.XGDP.56.FSGOV
3	Government Expenditure	Government expenditure per secondary student a...	-0.14	SE.XPD.SECO.PC.ZS
4	Government Expenditure	Government expenditure per tertiary student as...	-0.35	SE.XPD.TERT.PC.ZS
5	Gender Equality	Gross graduation ratio from first degree progr...	-0.16	SE.TER.CMPL.ZS
6	Gender Equality	Graduates from tertiary education, female (num...	0.20	SE.TER.GRAD.FE
7	Gender Equality	Gross enrolment ratio, tertiary, female (%)	0.49	SE.TER.ENRR.FE
8	Gender Equality	UIS: Percentage of population age 25+ with at ...	-0.16	UIS.EA.5T8.AG25T99.F
9	Labor Force Education	Labor force with basic education (% of total)	0.02	SL.TLF.BASC.ZS
10	Labor Force Education	Labor force with intermediate education (% of ...	-0.03	SL.TLF.INTM.ZS
11	Labor Force Education	Labor force with advanced education (% of total)	0.02	SL.TLF.ADVN.ZS
12	PISA Score	PISA: Mean performance on the reading scale	-0.09	LO.PISA.REA
13	PISA Score	PISA: Mean performance on the mathematics scale	-0.36	LO.PISA.MAT
14	PISA Score	PISA: Mean performance on the science scale	-0.26	LO.PISA.SCI

Round off the KG

Populate the relations between Employment rate and Indicator with corresponding Correlation Coef.

```
for indicator_code, correlation_df in zip(corr_df.indicator_code, corr_df.correlation_coef):
    query = """
        MATCH (n:`Educational Indicator` {indicator_code: $indicator_code})
        WITH n
        MATCH (n2:`Employment Outcome`)
        MERGE (n)-[r3:correlates_with {correlation_coef: $correlation_coef}]->(n2)
    """
    result = session.run(query, indicator_code=indicator_code, correlation_coef=correlation_df)
```



Reflections & Challenges

Are KGs a good solution to our research question? To what extent do KGs benefit our research question?

- A Hierarchical Structure of Indicators
- Lack of Foreign Key => Lack of topological graph representation
  - Not many useful queries could be made that benefit our research
- Using KGs in our task vs. Traditional Relational Data Approach: Correlation Analysis

Socioeconomic Status:

After filtering out Europeans countries, only two countries fall in "Middle" category

Lack of samples makes us unable to carry out relevant study.

Future research may involve binning of the country to stratify the country socioeconomic status.

Take-home Message

- ✓ Among the educational indicators we covered in our research, **Gross enrolment ratio of females in tertiary education** is most positively correlated with educational outcomes, while **Government Expenditure on Secondary Education as % of GDP** and **mean PISA score on Math Scale** are significantly negatively associated with employment outcome.
- ✓ Not enough Socioeconomic background data after filtering.
- ✓ It is critical to reflect on the optimal approach to the research questions.

TU/e