A graph-theoretical approach to fusing teachers' opinions about learning outcomes in higher education programmes

Carsten B. Guggedal

June 7, 2021

NTNU

Department of Engineering Cybernetics

Abstract

Higher education programmes can be modelled as graphs were nodes may represent courses, subjects, topics and sub-topics within a program, and where the edges represent potential relations among them. Such graphs may be used to represent intended learning paths through constructivistic interpretations like *"to learn A one needs to know first B"*, where A and B may be content units from the programs with different levels of granularity. These graphs are, however, designed by hand and are subject to personal interpretation. In this thesis we explore how to translate abstract connections of learning outcomes expressed through what are called Course Flow Matrices into more detailed connections using a graph-theoretical approach.

In detail, five methods that translate abstract connections are considered. The methods are analysed using graph theoretical tools such as centrality and connectivity. The centrality of the graph is evaluated with four different metrics, namely degree, PageRank, Katz and closeness centrality. The metrics are used to determine the importance, influence and closeness of the subjects in a course. The connectivity of a graph can identify if the translation result in inconsistencies or overlap in the programme.

This thesis investigates the benefits and drawbacks of using the individual methods for translating abstract connections, in terms of how well the connection represent the actual knowledge flow and how the graph properties change for the subjects.

By comparing the different methods on data from cybernetics-related programs, it is discovered that the method that connect to the nodes with the highest indegree yield the best results, with best an opportune metric that combines pedagogical interpretability with readability of the end results. If however one wants to maximize preserving some specific mathematical properties of the graphs possessed before the translation process (e.g., some specific type of connectivity), then the recommended method is that of connecting to nodes with no outgoing edges.

Sammendrag

Et høyere utdanningsprogram kan bli modellert som en graf hvor noder kan representere fag, emner, teamer eller sub-temaer innen et program, og hvor kantene representerer potensielle relasjoner mellom dem. Slike grafer kan bli brukt til å representere ment lærings flyt gjennom konstruktivistiske tolkninger som *"for å lære A trenger man å kunne B først"*, hvor A og B kan være en form for kunnskap fra programmer med forskjellig granulitet. Disse grafene er, derimot, designet for hånd og kan bli tolket personlig. I denne avhandlingen utforsker vi hvordan en kan tolke en abstrakt forbindelse av lærings flyt, uttrykt gjennom det som vi kaller Course Flow Matrices til mer detaljerte forbindindelser ved bruk av graf-teoretisk tilnærming.

I detalj, fem metoder som omformer en abstrakt forbindelse blir vurdert. Metodene blir analysert ved hjelp av graf teoretiske verktøy som sentralitet og konnektivitet. Sentraliteten av en graf blir evaluert ved hjelp av fire målinger, nemlig degree, PageRank, Katz og closeness. Disse målingene kan bli brukt til å bestemme viktigheten, innflytelsen og nærheten til et emne i ett fag. Konnektiviteten til en graf kan hjelpe til å identifisere om omformingen resulterer i inkonsekvenser eller overlapp i programmet.

Denne avhandlingen undersøker fordelene og ulempene ved de individuelle metodene for omforminger av abstrakte forbindelser, ved å se på hvor bra forbindelsene representerer den faktiske kunnskapsflyten og hvordan graf atributtene forandrer seg for hvert emne.

Ved å sammenligne de ulike metodene på data fra kybernetiske program, blir det konkludert med at metoden som lager forbindelser til noder med høyest indegree gir det beste resultatet, med en begrunnelse i både pedagogisk tolkningsevne og forståelse av sluttresultatet. Hvis, til formodning, en vil maksimere det å bevare noen matematiske attributer til grafen før omformings prosessen (i.e noen form for konnektivitet), så er den anbefalte metoden å lage forbindelser til noder som ikke har en kant utover.

Contents

1 Introduction

Graph theory is the study of graphs and the concept's origin date back to 1741. Graphs and graph theory can be a useful tool to help visualize and interpret how systems are connected. A graph can be used to model objects and the relations between them. The application of graphs include physical, theoretical, sociological and information systems. A common use of graph representation is within social media, where users can be modelled as nodes and the relation between them as edges [8]. The current application of graph theory within education and curriculum is however limited. In higher education, citations in scientific papers can be modelled as a network and represented as a graph [9].

Higher education programmes consist of a set of courses spread throughout semesters. Courses taught later on in a programme usually rely on knowledge obtained in earlier courses. The education programme can be represented as a directed graph where nodes represent a concept, topic or a course and the edge between them indicate the knowledge flow. A quantitative analysis of curriculum coherence has been conducted by Varagnolo et al. [11]. Such an analysis can help uncover inconsistencies, overlap and discrepancies in the education programme [16].

The different courses within a programme can be modelled as individual graphs were nodes represent a concept, topic or fact. The knowledge flow within the course is usually clear as the teacher knows what is taught and how it relates. A study conducted by the supervisors has shown that the student's opinion and teacher's opinion about a course may differ [11]. The knowledge flow between the courses may, however, not be as clear. Teachers may have different opinions on how the subjects between their courses relate. Is there a way to find the connections between the courses using graph theory and how does it relate to real world? This thesis seeks to answer that question.

1.1 Problem Description

A study programme in higher education consist of a set of courses which cover a range of different topics. Any programme can be described as a set of Knowledge Components (KCs) and the relation between them. A KC is defined as "*an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks*" [12]. In practise it translates to anything a student can learn, which can be in the form of practise, facts, concepts, subjects and procedures.

Most courses in a study programme have prerequisite knowledge required for the course. Another interpretation is that KCs are needed in order to learn other KCs. A course can in this way be represented as a directed graph of KCs. An approach for modelling a programme is using Course Flow Matrices (CFMs) as denoted by Wengle, Knorn and Varagnolo [16].

A CFM is built using various information from the teacher, which includes a course summary that lists the developed and required KCs, the Intended Learning Outcome (ILOs) and Teaching and Learning Activities (TLAs). The teacher also provides a Knowledge Component Matrix (KCM), which describes the relation between the developed KCs and the prerequisite KCs. Lastly, information about the developed KCs indicating which knowledge level student's should have for each KC [16]. A single course is typically represented by one CFM (or two complementing ones, as described in Section 3.4).

There exist several challenges regarding modelling higher education programmes. For an entire study programme there are tens of CFMs and hundreds if not thousands of KCs. Different teachers may construct the CFMs for different courses in the study program but how well does the knowledge flow translate between them? The different KCs within the CFMs are given on different forms of granularity or detail. Consider the two courses Mathematics and Chemistry, where Mathematics is a prerequisite for Chemistry. The Chemistry teacher may not know exactly which concepts are taught within Mathematics but knows that some of them are needed. Both courses consist of several KCs, but is every KC in Mathematics a prerequisite for Chemistry? Perhaps only half of the KCs in Mathematics are needed in order to learn Chemistry.

Assume then that the Chemistry teacher says in her CFM that her course needs 'Mathematics'. This means that the students will have a rather coarse representation of what they actually need to know. Assume moreover that the database where a university stores all its CFMs there exist the CFM that details the course 'Mathematics'. This thesis investigates different methods to expand the CFM of Chemistry by adding to it, in an intelligent way, the CFM of Mathematics, by finding a purely data driven way to derive the connections among the KCs within Mathematics and the KCs of Chemistry. In other words, the thesis' goal is to help students get a higher detail of which KCs they need to learn by automatically combining and merging CFMs given by different teachers that may not have even ever met and discussed about their courses.

We remark from the beginning that an important challenge arises due to the fact that CFMs and knowledge flow are subject to personal interpretation. As a consequence, a potential drawback of modelling study programmes using CFMs, is that the CFMs have to be constructed by hand. A course may be taught by several teachers and if they all were to construct a CFM of the course, the matrices may not be equal. Different teachers can have a different view on the course content and how the subjects taught within it relates. The student's opinions might also differ from the teachers as shown by Varagnolo et al. [14] [11]. These issues are not tackled in this thesis and need to be addressed in the future. In other words, this thesis shall be intended as a first step towards a computer assisted strategy for merging CFMs that may, if desired by the users, account also for personal opinions.

1.2 The role of this thesis from a graphical perspective

As mentioned above, this thesis investigates different methods of interpreting and fusing teachers opinions about learning in higher education. The CFMs provided by teachers may indicate a connection between KCs from different courses but not necessarily to what extent.

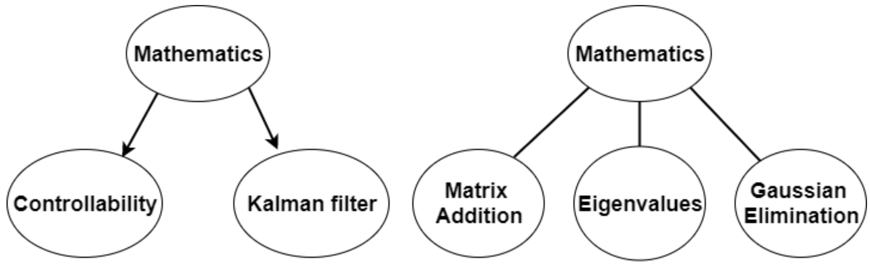


Figure 1: An example figure of how KCs from different courses connect. The graph on the left show that 'Mathematics' is needed in order to learn 'Controllability' and 'Kalman filter'. The figure on the right show some of the subjects taught in the 'Mathematics' course.

The objective of the thesis is to investigate how to translate more abstract KCs into smaller and more defined KCs. The smaller KCs can be thought of as a community of categorized KCs described only by an abstract KC. To make a practical example, consider figure 1, that shows that the KC 'Mathematics' can be translated into the smaller KCs 'Matrix addition', 'Eigenvalues' and 'Gaussian Elimination'. The CFMs are used to model the courses as graphs where nodes represent KCs and the dependencies between the KCs are represented by directed edges. Because the application is made for knowledge networks in higher education, the questions we seek to answer are

- how well do the translated KCs reflect the course dependencies?
- how do the graph properties change with the different algorithms?
- which graph properties should the translation be based on?

We are trying to answer these question to see if there is a logical way to connect KCs given an abstract connection. In detail, we want to look at different graph theory concepts such as centrality and connectivity to determine what happens to the graph and what it means from a pedagogical point of view. If the graph becomes disconnected it means that there are KCs that have no connection. On the other hand, if the graph becomes strongly connected it indicates that there are inconsistencies in the programme. The centrality indexes indicate what happens to the individual KCs. For instance, if the Closeness centrality index decreases, it would mean that the KC is less close to the other KCs. If the PageRank centrality decrease it would mean the KC is less influential than it was before. Looking at different centrality metrics and connectivity we want to learn which algorithms provide the most useful result and see which properties the translation should be based upon.

1.3 FACE-IT

Fostering Awareness on program Contents in higher Education using IT tools (FACE-IT) is a project funded by Erasmus and Strategic Partnership. The main focus of the project is to quickly and efficiently get and maintain holistic maps of program contents. Such a tool would be of practical use to teachers, program boards and students alike, since they often lack a holistic viewpoint. The holistic maps are represented as weighted directed graphs which are called Domain Model Graphs (DMGs), were nodes represent Knowledge Components (KCs) and the relation between them. A course-wide DMG (CW-DMG) depicts the knowledge flow through a course and is constructed by the teacher. All the teachers within a study program may combine their CW-DMG and together form a program-wide DMG (PW-DMG). For a study program it usually follows that some courses are prerequisites for other courses later on in the program and thus the PW-DMG depicts a knowledge flow the students would usually follow through the program. The PW-DMGs can be of use to students, universities and industry to detect any discrepancies in the program, ease communication between universities and ease the implementation of the Bologna agreements.

1.4 Motivation

In this paper we will investigate how higher education programs can be modelled as graphs and how different courses and subjects in the education program are connected. Courses and subjects are represented as graphs consisting of *super nodes* and *atomic nodes*. A *super node* is a node that can be split into smaller *atomic nodes*, see section 2.5 for the full definition. A *super node* can appear in a graph due to lack of information regarding how certain subjects are connected. This thesis focuses on how the *super nodes* and their *super edges* are translated into *atomic* ones using different methods. These methods are then evaluated using existing graph theory analysis tools such as connectivity and centrality to analyse the resulting graph. It is important that the resulting graph properties remain close to the original graph properties, because the graphs have a pedagogical interpretation. For instance, a graph should not become less connected after translating the *super nodes*, as it would mean that there is less coherence between the subjects in the course. Furthermore, a decrease in the PageRank centrality for some node would result in the subject being less influential in the course. Analyzing the different centrality indexes and connectivity is therefore useful to determine the validity of the result and the usefulness of the method. Hopefully, the methods developed and their results can be used as a tool to aid in the structuring of courses, semesters and study programs.

2 Theory

In this chapter, the relevant theoretical background and information will be given. First off, the fundamentals about graphs, such as structure and other properties, will be given. Afterwards, the concepts of graph connectivity and centrality will be defined along with the application for higher-education.

2.1 The Graph

A graph is a mathematical structure used to model objects and the relation between them. The objects within a graph are most often called nodes (can also be called vertices or points) and the relation between them are called edges (can also be called links or lines). The mathematical definition of a graph G is given by

$$G = (V, E) \tag{1}$$

where V denotes a set of nodes and E denotes a set of edges that define the connection between the nodes in V.

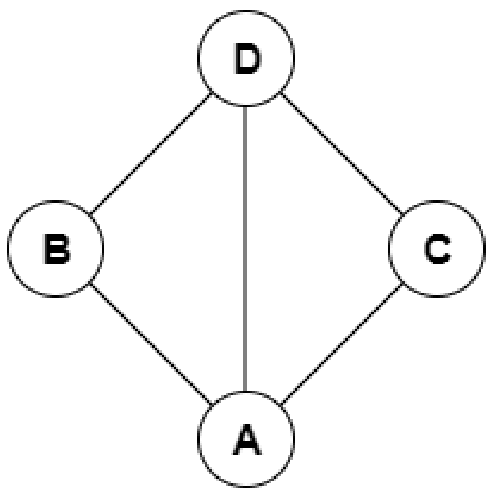


Figure 2: The Diamond graph. The graph is undirected and has four nodes and five edges.

More specifically, for a graph G with n nodes and m edges, the sets are given by $V = \{v_1, ..., v_n\}$ and $E = \{e_1, ..., e_m\}$. The set of edges E for a graph is defined by

$$E \subseteq \{\{u, v\} \mid u, v \in V \text{ and } u \neq v\} \tag{2}$$

where $\{u, v\}$ is an edge joining the two nodes u and v. An edge can either be undirected or directed. If the edge is directed, the order of the nodes u

and v in the edge $\{u, v\}$ indicate that the edge is directed from u to v. The set of edges E for a directed graph is defined by

$$E \subseteq \{\{u, v\} \mid u, v \in V^2 \text{ and } u \neq v\} \tag{3}$$

A graph is said to be directed if the edges are directed. A graph can be weighted, in which case each edge $e \in E$ has a weight $w > 0$ associated with it.

A trail p_{uv} is a set of edges $E_s \subseteq E$ containing the edges that connect nodes u and v. A path is a trail where all nodes connected by the set of edges E_s are unique. A shortest path is a path that minimizes the number of edges in E_s. If a graph is weighted, a shortest path is a path that minimizes the total weight of the edges within the path.

2.2 Connectivity

A graph is said to be connected if every pair of nodes is connected. Any two nodes u and v are said to be connected if there exists a path from u to v. If there does not exist a path between them, the nodes are disconnected [5]. If an undirected graph has any disconnected nodes, the graph is disconnected.

For directed graphs there are three subcategories of connected graphs:

- a graph is strongly connected if every pair of nodes are connected in both directions.
- a graph is semiconnected if every pair of nodes are connected in either direction.
- a graph is weakly connected if replacing every directed edge with an undirected edge yields a connected graph.

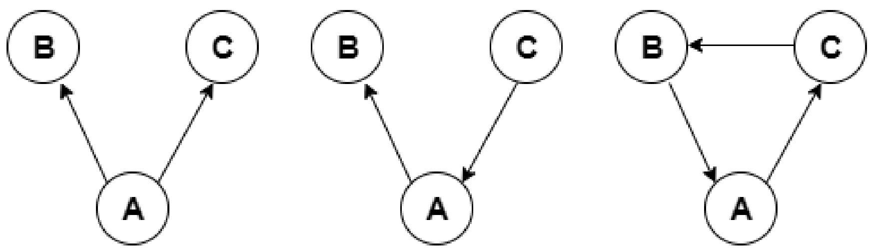


Figure 3: An illustration of the different subcategories of connectivity for a directed graph. Left figure: weakly connected. Middle figure: semiconnected. Right figure: strongly connected.

A node u is a neighbor of node v if $(u, v) \in E$. For directed graphs, the node u is an in-neighbor of node v and the node v is a out-neighbor of node u if $(u, v) \in E$. Nodes that are in-neighbors are often called predecessors and nodes that are out-neighbors are often called successors.

A graph G is a Directed Acyclic Graph (DAG) if G is directed and there exist no cycles in G.

2.3 Adjacency Matrix

The adjacency matrix A of a graph G is a square matrix of size $n \times n$. The adjacency matrix is denoted by

$$A = (a_{i,j}), \quad A \in \mathbb{R}^{n \times n} \tag{4}$$

where $n = |V|$ is the number of nodes in G and $a_{i,j}$ is the weight for the edge $(i, j) \in E$. For an unweighted graph the element $a_{i,j}$ is 1 if connected, i.e $(i, j) \in E$ or 0 if not [1]. For a directed graph, the adjacency matrix A is often asymmetric. The adjacency matrix of an empty graph is a zero matrix.

2.4 Centrality

In a graph some nodes may be more important than other nodes in one way or another. The importance of a node depends on what property the graph is evaluated on. This could be how many edges a node has, if it is a part of several shortest paths, etc. The importance of a node is measured by different "centrality indexes" which calculates the importance based on a certain index. In this paper six such centrality indexes will be considered and evaluated in order to analyze the CW-DMGs. The different centrality indexes have a pedagogical interpretation [14] and seeing how the algorithms affect the different indexes can provide a better view of which ones perform best.

Degree centrality is simply the sum of all weights of the edges connected to a particular node, or in the case of unweighted edges it is the sum of edges [5]. For a node v the degree centrality is defined as

$$C(v) = deg(v) \tag{5}$$

For a directed graph, degree centrality can be split into two smaller measures, indegree and outdegree. The indegree of a node v is equal to the degree centrality when only taking edges directed to the node v into account. The outdegree of a node v is equal to the degree centrality of a node v when only taking edges directed from the node v into account. From a pedagogical point of view the degree centrality indicate how many courses a particular concept is relevant for or how many course are connected to a particular course. The metric is sensitive to the weighting of the connection as described in [14].

Closeness centrality ,[6], is the average length of the shortest path between a specific node and all other nodes in the same graph. For a node u the closeness centrality is defined as

$$C(u) = \frac{N-1}{\sum_{v=1}^{N-1} d(v,u)} \tag{6}$$

where N is the number of nodes and $d(v,u)$ is the shortest path distance between nodes v and u. For a directed graph, the shortest path distance is computed as the incoming distance to the node u. An improved improved method was introduced by Wesserman and Faust for graphs with more than one connected component

$$C(u) = \frac{n-1}{N-1} \frac{n-1}{\sum_{v=1}^{N-1} d(v,u)} \tag{7}$$

where n is the number of reachable nodes and N is the total number of nodes. The method takes into account the fraction of which nodes are reachable [7]. The closeness of a particular node roughly express how close that node is to the other nodes in the graph. The practical meaning is how far a concept/course is from other concepts/courses. Concepts used over the entire program in several course have a higher index value, concepts only taken up in courses close in the program/over a shorter period of time, have a lower index.

Eigenvector centrality (eigencentrality) assignes a score to each node in the graph. Higher scoring nodes contribute more to connections than low scoring nodes (with same weight) [13]. For directed graphs eigenvector centrality can yield negative values if the graph is not strongly connected, see Section 2.2. The definition of eigenvector centrality is given by

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t \tag{8}$$

where $M(v)$ is a set of neighbors of v and λ is some constant. The equation can be rearranged and written as

$$Ax = \lambda x \tag{9}$$

where $A = (a_{v,t})$ is the adjacency matrix. For a directed graph the adjacency matrix is in general asymmetric and there are two leading eigenvectors. The "left" eigenvector centrality corresponds to the in-edges in the graph and the "right" eigenvector centrality corresponds to the out-edges in the graph. Eigenvector centrality is a measure of how influential a given course/concept is within the program as a high score indicates that the concept is relevant in many other courses throughout the program.

PageRank centrality is very similar to eigenvector centrality but uses a different metric when assigning scores.

$$x_i = \sum_j a_{ji} \frac{x_j}{L(j)} + \frac{1}{N} \tag{10}$$

where

$$L(j) = \sum_i a_{ji} \tag{11}$$

is the number of neighbors to node j. The practical interpretation and usefulness of the result is expected to be similar to eigenvector centrality as the methods are calculated in a similar fashion.

Katz centrality [10] is similar to eigenvector and PageRank centrality and it is calculated based on the centrality of its neighbors. A node with many edges will get a high score but a node with a low number of edges may also have a high score, if its neighbors are high scoring. The method also gives each node a small amount of centrality for "free", in this way the algorithm handles directed graphs better than eigenvector centrality. Katz centrality of a node v is calculated using the formula

$$C_K(v_i) = \sum_{j=1} A_{ji} C_K(v_j) + \tag{12}$$

where A is the adjacency matrix, is called the damping factor and is a bias constant which controls the initial centrality. The damping factor satisfies the equation

$$< \frac{1}{_{max}} \tag{13}$$

where $_{max}$ is the largest eigenvalue of the adjacency matrix A [2]. Due to the similarities with eigenvector and pagerank centrality, the results are expected to be similarly useful.

Betweenness centrality uses the concept of betweenness, which looks at how often/how many times a certain node act as a 'bridge' between any two other nodes in the shortest path. This metric is expected to be low in most graphs because the concepts are exclusively connected to courses and vice versa. From a pedagogical point of view the metric indicate which concepts have the highest learning outcome based on the other courses or concepts to come. As a consequence of the metric being low in most graphs, the metric does not provide as useful results as other centrality metrics. Betweenness centrality is calculated using the following equation

$$g(v) = \sum_{s,t \in V} \frac{(s,t|v)}{(s,t)} \tag{14}$$

where (s,t) is the number of shortest paths from node s to node t, whereas $(s,t|v)$ is the number of those same paths that go through node v [3].

2.5 Notation

The following notations will be used throughout this paper to differentiate between nodes and edges which can be translated into different nodes and edges, and those that cannot. Nodes and edges which can be translated will be called *super nodes* and *super edges*, while those than cannot will be called *atomic nodes* and *atomic edges*. The atomic nodes are denoted with small letters (and indices where possible), atomic edges are denoted in a similar fashion. Furthermore, the sets of all atomic nodes and atomic edges are denoted without any subscript. For instance, the two atomic nodes v_i and v_j are elements in the set V and the atomic edge e_{ij} between these two atomic nodes is an element in the set E. The atomic edge e_{ij} may also be denoted as (v_i, v_j).

A super node is a set consisting of atomic nodes and is denoted with capital letters V_i, and the set of all super nodes is denoted V. We also assume that no atomic node is an element of two super nodes, so if we have T super nodes then

$$\bigcap_{i=1}^{T} V_i = \emptyset \tag{15}$$

In a similar fashion, the set of atomic edges between the atomic nodes in V_i is denoted as E_i, i.e.

$$E_i = \{(v_i, v_j) \in E \mid i \neq j, v_i \in V_i \text{ and } v_j \in V_i\} \tag{16}$$

A super node can therefore be interpreted as a *subgraph* constructed as

$$G_i(V_i, E_i) \tag{17}$$

Furthermore, the set of atomic edges between two super nodes V_l and V_k is denoted as E_{lk} with

$$E_{lk} = \{(v_i, v_j) \in E \mid v_i \in V_l, v_j \in V_k\} \tag{18}$$

Consequently it follows that $E_i \subseteq E$ and $E_{ij} \subseteq E$.

A super edge is a set consisting of atomic edges and is defined as $\bar{E}$. A super edge is connected to at least one super node. If a super edge is connected to an atomic node v_i and a super node V_j, all atomic edges e_i in $\bar{E}$ are connected to the atomic node v_i. The set of atomic nodes in V_j that the atomic edges e_i connect to is undefined.

3 Algorithms and Data

In this section we will cover the different methods/algorithms used for translating super nodes into atomic nodes. There are several different ways to do this translation and for each method the graph properties will change. By evaluating the different properties of the resulting graph, the validity and usefulness of the algorithm can be determined. From a pedagogical point of view, the methods should preserve some graph properties because it has a practical interpretation. A simplified example is included to illustrate how CMFs are represented as graphs. The example also illustrates how a super node is translated into atomic ones using a basic method. The results are then analysed, which forms the basis in the same manner as the results described in section 4.

3.1 Algorithm Classification

In this thesis, five algorithms are developed to translate abstract KCs into detailed connections. Each method chooses the atomic nodes to connect to based on a graph property:

The first method is a naive method that connects to every single atomic node in the super node. The method is very basic and will serve as a comparison against the other methods.

The second method connects to the atomic nodes which have the highest in-degree. Nodes that have a high in-degree can be interpreted as complex subjects that require a lot of previous knowledge. If there are multiple nodes that have the highest in-degree, an atomic edge will be added to every one. The number of atomic edges is expected to be low for this method.

The third method connects to the atomic nodes which have no outgoing edges. Nodes that have no outgoing edges can be interpreted as the learning outcome or learning goals of the course. The method is expected to yield a low to medium amount of atomic edges.

The fourth method connects to the atomic nodes that have the highest degree. The nodes that have the highest degree will typically be central subjects which require previous knowledge but also is a requirement for other subjects. As with the second method, the method is expected to yield a small amount of atomic edges.

The fifth and final method connects to the atomic nodes that have both incoming and outgoing edges. These nodes require previous knowledge and is a requirement for other subjects as with the fourth method, but the amount of connections is of no concern. The number of atomic edges is expected to be medium high.

The methods are expected to yield a different amount of atomic edges and connect to different atomic nodes.

3.2 Property Preservation

In order to analyse the validity and usefulness the different methods described in Section 3.1, an analysis of how different centrality indexes change is made. For a method to be valid, it should preserve some graph property. This is the result of the graph properties having a pedagogical interpretation. Subjects that are important before the translation should not become less important. A graph should also not become less connected as a result of the translation. Depending on the method, certain centrality indices may change more than others. The result could be used to determine which method to choose for the translation based on the need to preserve a given property. The centrality indexes evaluated in this thesis are degree centrality, PageRank centrality, Katz centrality and closeness centrality, described in Section 2.4. The eigenvector and betweenness centrality measures will not be evaluated for the methods because they are expected to yield poor results. Eigenvector centrality is based on the centrality of their neighbors and is suitable for undirected graphs or strongly connected directed graphs. The graphs used throughout this thesis are directed but not expected to be strongly connected and the metric will therefor yield poor results. PageRank and Katz centrality is evaluated instead, both of which are variants of eigenvector centrality. Beetweenness centrality has also been omitted as the results are expected to be low or zero. This is the result of the typical graph structure for a knowledge graph, where subjects are exclusive to a course. For more information on the different centrality indexes and their interpretation, see Section 2.4.

Additionally, the graph connectivity should preferably not change, but at the very least not decrease. If the graph becomes disconnected, there are subjects which are not connected or have no overlap. On the other hand if the graph become strongly connected, there are inconsistencies in the problem.

Seeing how the different properties change using a particular method is useful as it can help determine which method to choose based on which properties are desired in a certain graph.

Remark (on the parameters of the Katz and PageRank centrality indexes) we note that both Katz and PageRank centrality utilize the damping parameter . The initial centrality given to nodes is calculated using in PageRank centrality (see eq. 10), but Katz centrality use the parameter (see eq. 12). In this thesis the damping factor for PageRank centrality is set to $= 0.85$, following the justifications of the choice made by [4]. The damping factor for Katz centrality is set to $= 0.1$, which was the default value for the method used for evaluation. The initial centrality is set to 1 for simplicity.

3.3 The dataset used to tune the algorithms

For the evaluation of the different merging algorithms two courses are modelled as graphs. Each course have two graphs associated with it. The first graph describes the different subjects and topic thought within the course. This graph is represented as a tree with each child being a sub topic of its parent. The other graph describes what subject or topic is needed to learn which other subjects. This graph may contain nodes from its own subjects and subjects from other courses. If a node is not present in the graph it means that nothing is needed in order to learn that subject.The first graph is known as the specification graph and the latter as the learning graph. The two courses used throughout this thesis is a representation of the Systems Theory course and Algebra course. The Systems Theory graph will have red nodes and edges while the Algebra course will have blue nodes and edges. The translated atomic edges are represented with dashed black edges.

3.3.1 Systems Theory

Steffi Knorn, a professor at the Technische Universität Berlin, compiled a CFM for the ‘Systems Theory’ course with the aid of Magnus Axelson-Fisk. The author modified the CFM by reducing the amount of KCs. The reason for this was to gain a smaller graph for better visualization and interpretation.

It is important to note that some of the KCs from this course depend on the rather generic concept ’Algebra’. This, as shown below, is actually a course by itself.

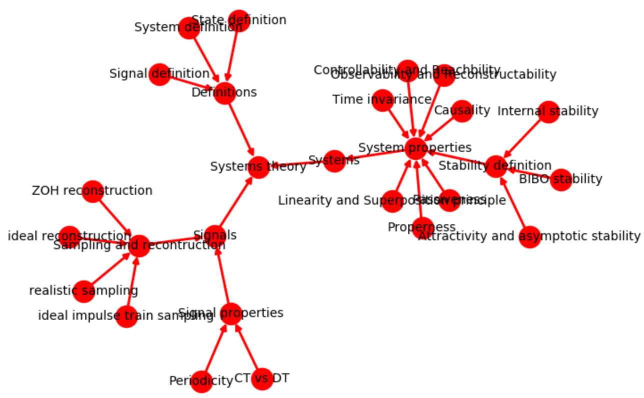


Figure 4: The specification graph for the Systems Theory course.

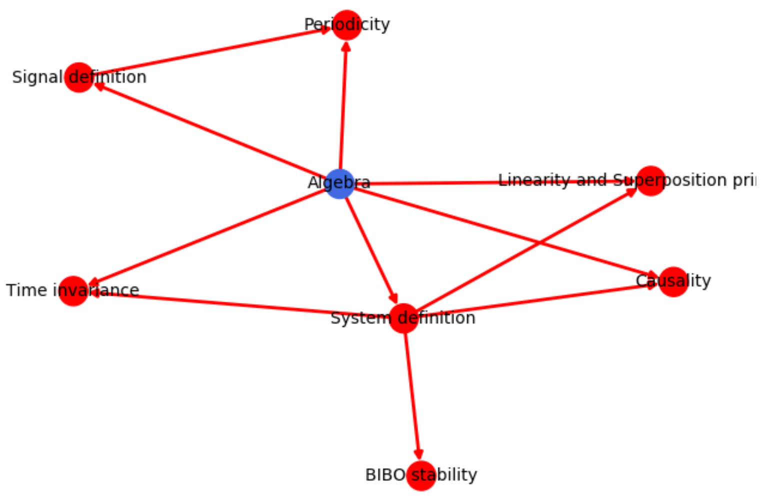


Figure 5: The learning graph for the Systems Theory course. The generic concept ‘Algebra’ is represented by the blue node.

3.3.2 Algebra

Following up the comment above, we exploit the fact that Algebra is a course by itself in higher education. We thus consider a CFM for the ‘Algebra’ course that was compiled by the author based on data provided for a similar course used by Emil Wengle in [15]. The CFM was designed to be of similar size as the CFM for ‘Systems Theory’.

We may then rephrase the purpose of the thesis based on the two CFMs above in these terms: expand the CFM of ’Systems Theory’ by expanding its ’Algebra’ node with the CFM of Algebra, and by adding an opportune set of edges that connect the super-set of nodes obtained in such a way.

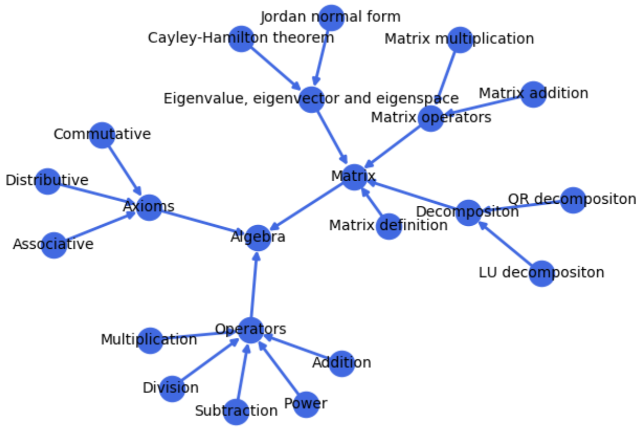


Figure 6: The specification unit graph for the Algebra course.

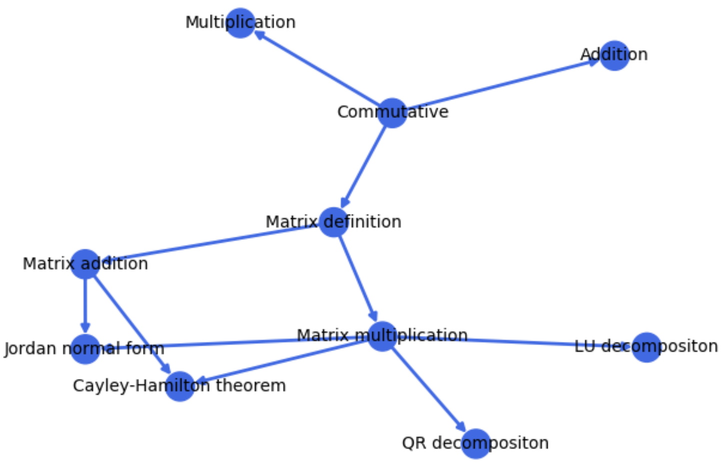


Figure 7: The learning graph for the Algebra course.

3.4 Illustrating the example above in a graphical way

This section generalizes and visualizes the problem stated above through an example that we perceive is clearly illustrating, in a generic way, how courses are represented as graphs. Assume a course to be represented using two graphs:

1. the first, that we may call the ***hierarchy*** of the contents of the course, containing the information about each KC taught within the course

through a tree structure (i.e., a graph where each child node should be intended as a smaller subject within its parent);

2. the second connecting what KC or courses are needed in order to be able to learn another given KC. For this we may call this graph the ***logical flow*** of the course, i.e., how one should proceed in a temporal way in her/his learning efforts.

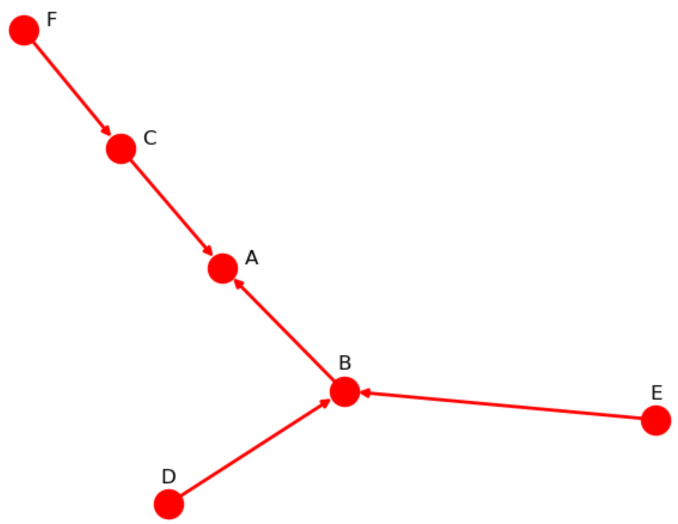


Figure 8: A simplified graph representing the course A and the subjects B-F taught within the course. The subjects B and C are direct subjects of course A, while D and E are smaller subjects of B.

Figure 8 illustrates what subjects are taught within course A. The course A and the subjects B, C are considered super nodes as they have smaller subjects contained within them. The subjects F, D and E have no smaller subject contained within them so they are considered atomic nodes.

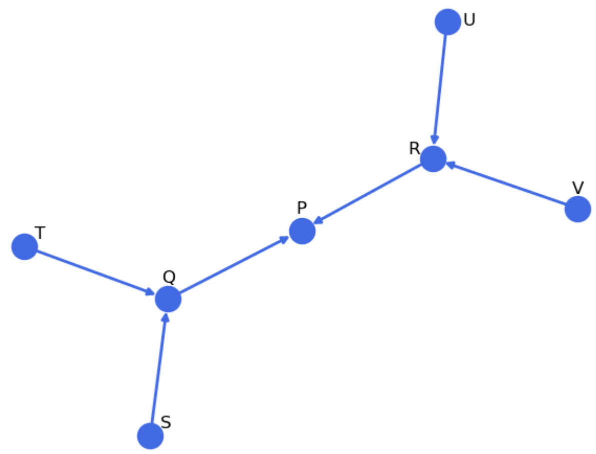


Figure 9: A simplified graph representing the course P. The course consist of the subjects Q and R, where both Q and R consist of smaller subjects.

Figure 9 illustrates the same as figure 8, but for course P. In this graph the course P and subjects Q and R are super nodes, and as a consequence the subjects T, S, U and V are atomic nodes. From a pedagogical point of view the atomic nodes are essentially everything that is taught in the course and therefore make up the entire course.

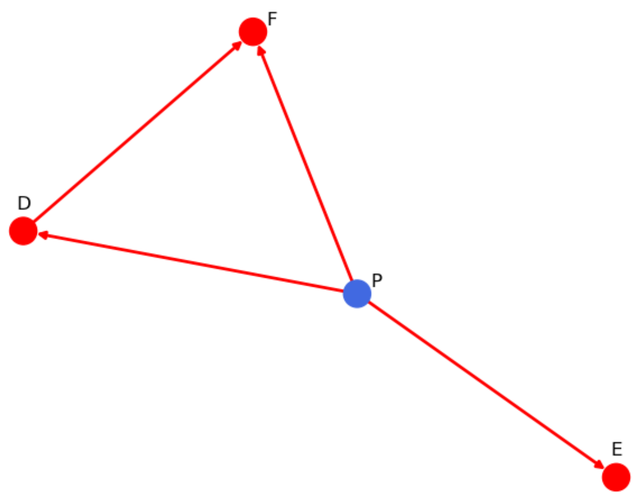


Figure 10: A simplified graph representing what is needed in order to be able to learn certain subjects in course A.

What subjects or courses are needed to learn a specific subject within course A is illustrated in Figure 10. Course P is connected to every node in the graph so a student must have had course P to be able to learn any other subject in the graph. Subject D, which is taught in course A, is needed to learn subject F in the same course.

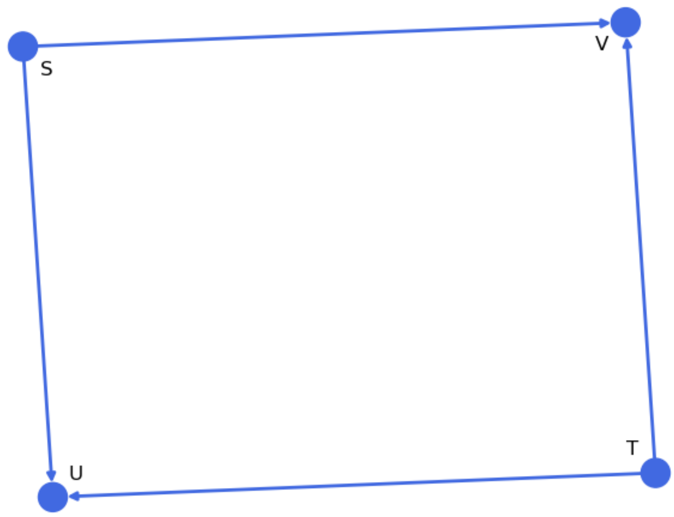


Figure 11: A simplified graph representing what is needed to learn the different subjects in course P. The subjects S and T taught within course P are needed in order to learn subjects U and V from the same course.

Figure 11 shows what subjects or courses are needed in order to learn subjects within course P. Both U and V are dependant on subjects S and T from course P. The subjects taught in course P therefore does not require the student to have taken any other courses.

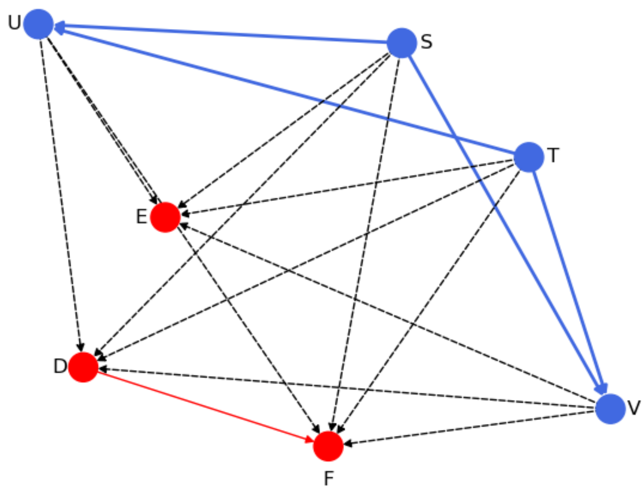


Figure 12: A graph representing the A course learning information shown in Figure 10, after replacing the super nodes and edges with atomic ones. The super nodes are replaced with all the atomic nodes contained within the subgraph of the specific node. Red nodes and edges represent course A and the blue nodes and edges represent course P. The dashed black edges are added atomic edges when replacing super nodes and edges.

From figure 10 it can be seen that the nodes within course A depend on the contents from the P course. The graph in Figure 12 illustrates the same learning graph from Figure 10 but the super nodes and edges have been replaced with atomic ones. For this example there was just a single super node, namely P, but for larger graphs there may exist several super nodes and edges. A graph may contain super nodes from itself, i.e subjects within the course or it may contain super nodes from other subjects or courses. Analysis of the merged graph is done using different centrality measures such as degree centrality, PageRank centrality, Katz centrality, etc. Eigenvector centrality and betweenness centrality are expected to yield poor results, but are included in this simplified example to verify this. The different centrality measures used have different pedagogical interpretations and it is therefore useful to look at each of them to see what happens to the course in different aspects.

3.4.1 Degree Centrality

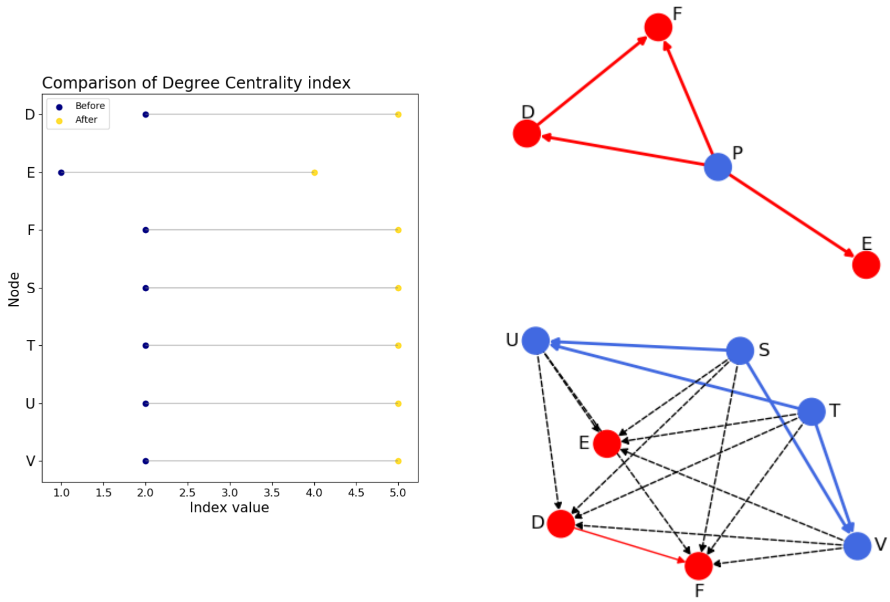


Figure 13: Degree centrality calculated on the learning graph of course A, before and after having replaced super nodes and edges.

Figure 13 show how the degree centrality change when replacing super nodes and edges with atomic ones. It can be seen that each nodes index increase with the same magnitude. The atomic nodes T, V, S and U that replace the super node P each has their magnitude increased by 3. This makes sense since node P had a degree centrality of 3. The other nodes also have their index value increased by 3, this is because one node and edge is removed and four new are added. The nodes that have their index increased are the nodes that have are connected to the super node, which in this case is every node. The centrality measure tells us about how important a given subject is in a given course. It is an indication on how many subject it is relevant for or how many other subjects are needed to be taught for this particular subject. This metric will remain the same or increase for all subjects. The metric can help identify subjects that previously were thought to not be important because they were connected to a super node, but after replacing the super node the index value increases and it can be seen that the subject is more important than previously thought.

3.4.2 Eigenvector Centrality

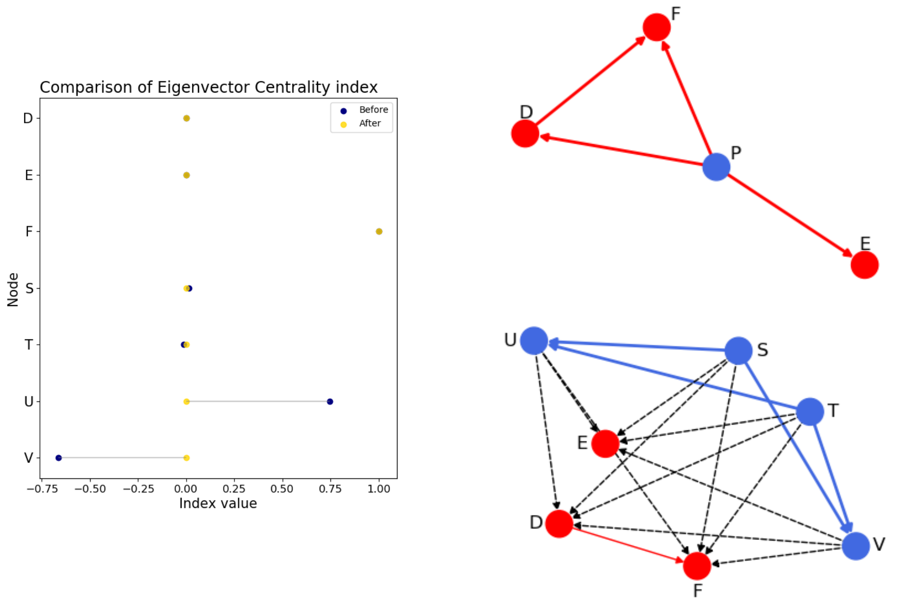


Figure 14: Eigenvector centrality calculated on the learning graph of course A, before and after replacing super nodes and edges.

Eigenvector centrality is calculated by assigning a relative score to each node with the idea that connections to high-scoring nodes provides a higher score. The results shown in Figure 14 show that for most nodes the value does not change or change minimally if so. The exceptions are node U which increases by 0.75 and the node V which goes from 0 to -0.75. The index value is calculated using the adjacency matrix of the graph, see eq. 9. Eq. 8 show that the value is calculated by taking the sum over all other nodes index value and multiplying it with the corresponding adjacency matrix element. This number will be zero if there is no incoming edge or 1 if there is an incoming edge, as a consequence most of the nodes in the graph will have an index value of zero since they have no incoming edges. Node E will have a zero index score because all its neighbors have a zero index score. The only node which has an index value larger than zero is node F. This is because it has an incoming edge from node D which has four incoming edges. The connection is therefore considered high scoring and will yield a high index score for node F. As explained in Section 2.4 the eigenvector centrality works poorly for directed graph, especially acyclic ones. The index value for node V before the merge is also observed to be negative which should not occur, and is an indication that the result is flawed. Eigenvector centrality can be interpreted as a measure of how influential a subject or course is, but because the result is flawed or zero it does not provide any meaningful insight. This problem is solved by the use of PageRank and Katz centrality.

3.4.3 PageRank Centrality

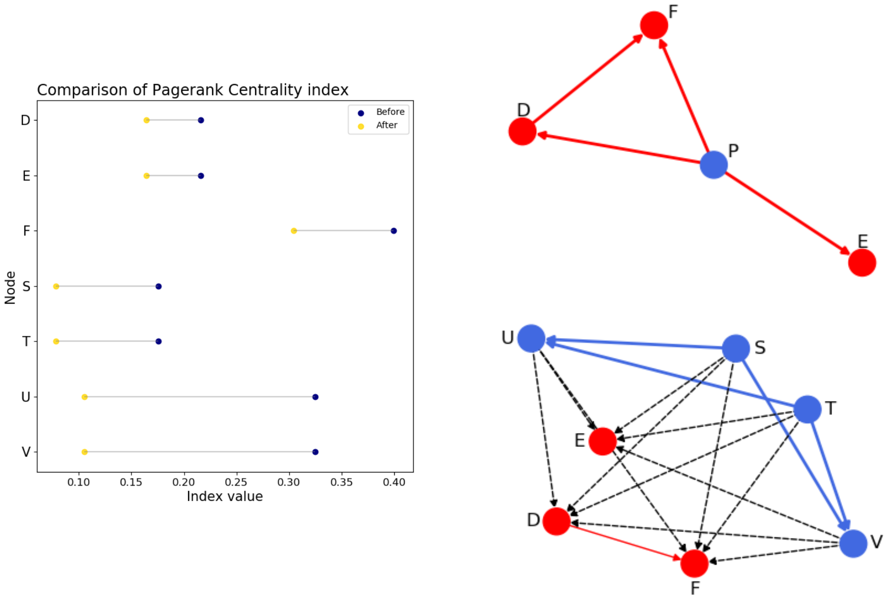


Figure 15: PageRank centrality calculated on the learning graph of course A, before and after replacing super nodes and edges. The index values are calculated using a damping factor of = 0:85.

From Figure 15 it can be seen that the PageRank centrality measure decreases for every node in the graph, but the magnitude differ. It is clear from the figure that the result from PageRank centrality differ from that of eigenvector centrality, since the method is able to account for link direction. PageRank also uses a scaling factor given by eq. 11 when calculating its index values. From Figure 15 there seems to appear only 4 distinct index values among the nodes. The inserted atomic nodes V, U, T and S each have an index value of 0:1 and 0:8 respectively. It is expected that these nodes should have the similar index value as they each have 3 outgoing edges to the same nodes. The difference comes from nodes U and V having incoming edges from nodes S and T. The nodes D and E share the same index value as well, but higher than the 4 other nodes. These nodes have 4 incoming edges which will contribute differently to the score than outgoing edges. Node D also have an outgoing edge to node F which will yield a higher score for node F which can be seen in Figure 15.

PageRank is an adaptation of the eigenvector centrality measure and can also be interpreted as how influential or important a subject is within the given course. The highest scoring node is node F which has 5 incoming edges, this subject is considered important as it requires 4 not as important subjects together with subject D, which is considered by subject F to be high scoring. A possible interpretation is that subject F is the learning outcome of course A. Subjects D and E also have a higher score and can be considered to be smaller learning goals of course A.

3.4.4 Katz Centrality

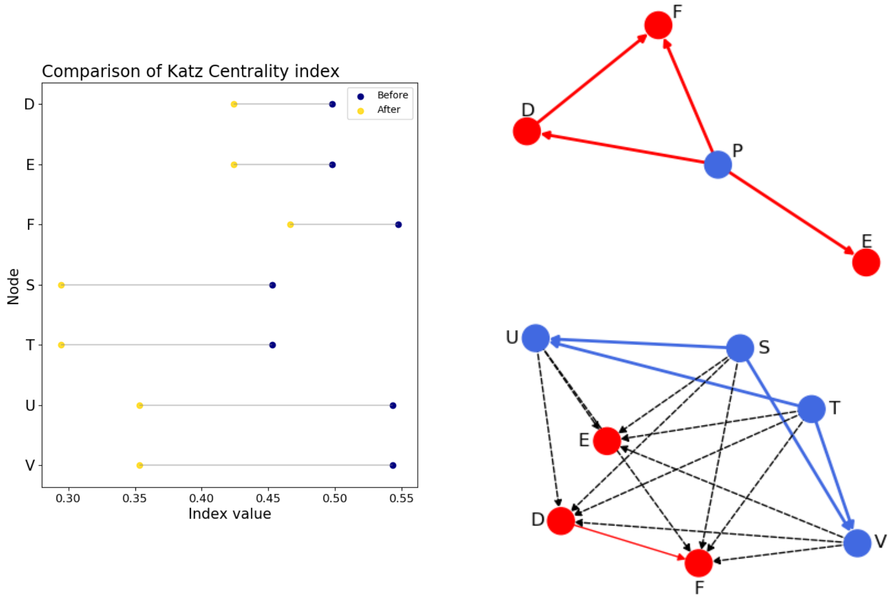


Figure 16: Katz centrality calculated on the learning graph ofcourse A, before and after replacing super nodes and edges. The index value is calculated using a damping factor of = 0:1 and an initial centrality of = 1:0

Figure 16 shows the Katz centrality index values before and after merge. Like PageRank centrality, it is an adaptation of eigenvector centrality but differs in the way the value is calculated. Katz centrality does not use a scaling factor like PageRank centrality, but like PageRank centrality it does give some centrality to each node by default. This is done with the parameter in eq. 12. Because the way both PageRank and Katz centrality are calculated it is expected that the result will be similar. By comparing Figure 15 and 16 it can be seen that the overall trend in the change of index values are the same. What differs is the magnitude at which they change which is caused by the scaling factor in PageRank centrality and the choices for the damping factor and initial centrality . The pedagogical interpretation of the result will be similar to that of PageRank and eigenvector centrality. Because the resulting index values are close to the values from PageRank centrality the same conclusion of the results can be made.

3.4.5 Closeness Centrality

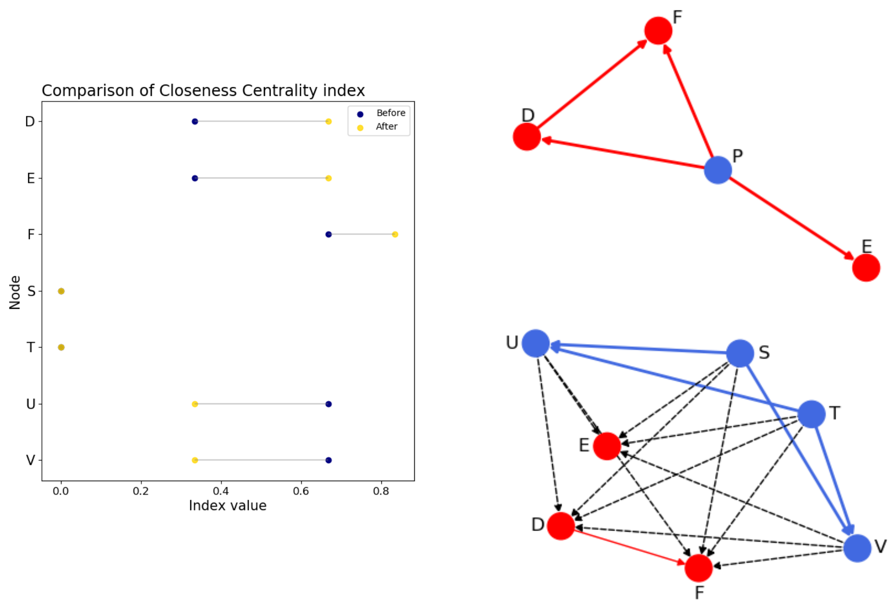


Figure 17: Closeness centrality calculated on the learning graph of course A, before and after replacing super nodes and edges.

Closeness centrality is a measure of the average shortest path to a node from the other nodes. Since the graph is directed the method calculates the incoming distance, i.e it follows the direction of the edges or the flow. From Figure 17 it can be seen that the index value of the nodes S and T are zero. This is because each of these nodes only have edges directed outwards and are therefore unreachable. The value for nodes D and E increase after having replaced the super nodes and edges, both of these nodes have four incoming edges after merge and their value is expected to be equal. As we are looking at incoming edges, outgoing edges are of little value to the respective nodes. The node F has five incoming edges after merge and two incoming edges before. The value is therefore expected to be high and to increase after the merge. An interpretation of this result is that subject F is close to a lot of other subjects, the same goes for subjects D and E. For subjects S and T, the index value tells us about how close they are to the other subjects, except that they have no incoming edges. It is possible to invert the direction of the edges, which would also reverse the result. The index value indicates which subjects are close to other subjects and vice versa. For larger and highly connected graphs, this metric will provide more useful results.

3.4.6 Betweenness Centrality

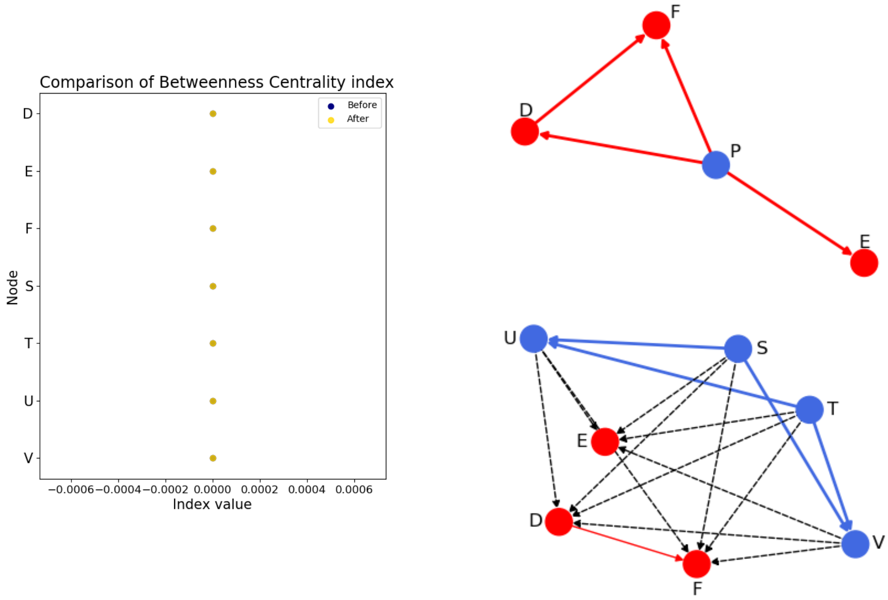


Figure 18: Betweenness centrality calculated on the learning graph of course A, before and after replacing super nodes and edges.

The results when calculating the betweenness centrality are shown in Figure 18. For all nodes the index value is zero, both before and after merge. This is a consequence of the structure of the graph. Beetweenness centrality is calculated by taking the number of shortest paths that goes through a certain node divided by all shortest paths. Because nodes U, V, T and S are unreachable they will get a value of zero. Nodes D, E and F also have a value of zero, as the shortest path between any two nodes in the graph is 1, consequently no shortest path will go through an intermediate node and all values will be zero. As for closeness centrality, for a larger and highly connected graph this metric will provide more insight. This metric is interpreted as which subjects are often needed in order to learn other subjects, but as a result of most graph structures in learning graphs, this metric will give limited insight.

4 Results on the field examples

In this chapter the results obtained from the algorithms described in section 3.1 are presented.

4.1 Naive merge

The first algorithm implemented is a basic method which simply replaces each super edge with atomic edges connected to every atomic node in the Algebra course. The algorithm is naive and will serve as a base algorithm for which the other algorithms will be compared against. The algorithm will find every atomic node from the specification graph and replace the super node, connecting the new nodes to each node that the super node was originally connected to. If any atomic node has any dependencies in the learning graph, then these will be present in the new graph given the dependant nodes are present.

Figure 19: The merged graph for the Systems Theory course using the naive method.

Figure 19 represents the same course as Figure 5 after replacing the 'Algebra' super node. Since each atomic node will have the same edges as the super node, the amount of edges will increase with a factor of $|V|(deg(Algebra) - 1)$. Additional edges may also come from dependencies caused by the atomic nodes. This is the case in Figure 19 and is observed by the blue edges.

4.1.1 Degree Centrality

Figure 20: Degree centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the naive method.

From Figure 20 it can be seen that all nodes except 'BIBO Stability' have their degree increased. This is because every node present in the original learning graph for the 'Systems Theory' course, shown in Figure 5, is connected to the 'Algebra' super node, except for 'BIBO stability'. The magnitude at which they increase comes from the number of atomic nodes replacing the super node. The atomic nodes share the same connections since they are all connected to the same nodes as the super node they replaced. Their respective degree may however be different because of dependencies among the atomic nodes, as shown in figure 7. As a consequence of replacing each super edge with atomic edges connected to every atomic node, the degree will increase by a large margin. Comparing the index values before replacing the super node and after, the difference from lowest scoring node to highest scoring node increases. Nodes from the Systems Theory course that are connected to the super node will have their index value increased by nine. Nodes from the Algebra course will have their index value increased by six, which is equal to the degree of the super node $deg(Algebra) = 6$.

4.1.2 PageRank Centrality

Figure 21: PageRank centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the naive method. The values are calculated using a damping factor of $= 0.85$.

PageRank centrality gives each node some centrality based on the damping factor . This results in each node having some influence by default when calculating the score. It is expected that the index value will decrease as a consequence of several new nodes and edges being added. The atomic nodes from the Algebra course have their index value decreased the most, and all the nodes have close to the same index value after merge. The index value for the nodes from the Systems Theory course also decrease but with a smaller magnitude than the nodes from the Algebra course, except for 'BIBO stability'. This node decreases more because it is not directly connected to the super node before the merge. Pagerank centrality offers insight into how influential a subject is within the course. As a consequence of the number of new nodes and edges being added from the merging algorithm, each subject can be interpreted as less influential than before. The subjects from the Algebra course are equally influential in the Systems Theory course. The nodes from the Systems Theory course are more influential than the subjects from the Algebra course.

4.1.3 Katz Centrality

Figure 22: Katz centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the naive method. The values are calculated using a damping factor of = 0:1 and an initial centrality of = 1:0.

Katz centrality is a variant of Eigenvector centrality and is also calculated based on the centrality of a nodes neighbors. The results shown in figure 22 are observed to be similar to those of PageRank centrality shown in Figure 21. The magnitude at which the index values change is different from PageRank centrality. Katz centrality uses the parameter as default centrality to each node, while PageRank centrality calculates the initial centrality based on the damping factor and the total number of nodes (see equation 10). From Figure 22 it can be seen that index values for every node is reduced after merge. This is also the result of the super edge being replaced with atomic edges connected to every atomic node in the Algebra course. The nodes 'Periodicity', 'Linearity and Superposition principle', 'Time invariance', 'Causality', 'Signal definition' and 'System definition' have their index value decreased but by a small margin compared to the other nodes. This is because they are present in the learning graph before the merge and will have a direct edge to every node from the Algebra course, see figure 5. 'BIBO stability' is also present in the original learning graph but the index value decreases more then the other nodes because does not have a direct link to the nodes from the Algebra course. The subjects present in the Systems Theory course learning graph before merge seems to be almost as influential after replacing the super node. The exception is 'BIBO stability' as mentioned before. The subjects replacing the super node becomes less influential than they were in the Algebra course. Subjects from the same course have almost the same influence as the other subjcets.

4.1.4 Closeness Centrality

Figure 23: Closeness centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the naive method.

Closeness centrality offers a different view on the graphs properties as the metric is flow based. Unlike the results from the link based methods, some nodes closeness centrality have their index value increased after merging as seen in Figure 23. The nodes that have their index value increased are the nodes that are present in the Systems Theory learning graph before merging, shown in figure 5. The reason for this is that the average distance to all other nodes, $\sum_{v=1}^{N-1} d(v;u)$, is decreased because there is a direct link to every node from the Algebra course. The node 'BIBO stability' does not increase as much as the other nodes because it is not directly connected to the nodes from Algebra, there are two edges between the nodes. For the atomic nodes replacing the super node, the index value decreases. For these nodes, no new nodes can be reached and the average shortest distance remain the same. The total number of nodes N in the graph increases which results in the index value decreasing. The node 'Commutative' have an index value of zero before and after merge, this is due to the being unreachable. Subjects already present in the Systems Theory learning graph will have a higher index after merge and will be closer to other subjects. The subjects that replace the super node becomes less close to other subjects than they were before. For the subject 'Commutative', there is no information on how close it is to the other subject, other than it only having outgoing edges.

4.2 Connect to nodes with highest indegree

This method looks at which atomic nodes has the highest indegree. This can be interpreted as replacing the super node with subjects that are complex or require the most previous knowledge. The method is expected to reduce the number of atomic edges replacing the super edges in comparison to the naive algorithm described in Section 4.1, unless every atomic node has the same indegree.

Figure 24: The merged graph for the Systems Theory course, using the method that connect to the nodes with the highest indegree.

From Figure 24 it can be seen that the super edges translates into atomic edges connected to the two atomic nodes 'Jordan normal form' and 'Cayley-Hamilton theorem'. The atomic nodes have an indegree of value 2, calculated on the corresponding learning graph shown in Figure 7. Every other atomic node has an indegree of either 0 or 1. The algorithm reduces the number of translated atomic edges compared to the naive method.

4.2.1 Degree Centrality

Figure 25: Degree centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest indegree.

The method reduces the number of atomic nodes which the translated atomic edges connect to, compared to the naive algorithm described in Section 4.1. Therefore the degree centrality is also expected to increase by a smaller amount. Figure 25 shows that the in degree is increased by 1 or stay the same for most nodes. The nodes that have their index value increased by one, were connected to the super node and will therefore have their degree index increased by one since two atomic nodes replace the super node. The node 'BIBO stability' is again unaffected because the node was not connected to the super node. The two atomic nodes 'Jordan normal form' and 'Cayley-Hamilton theorem' have their index increased by 6. This is a consequence of the super node having a degree $deg(Algebra) = 6$, see Figure 7. The other atomic nodes from the Algebra course have the same index value as no new edges are connected to them.

4.2.2 PageRank Centrality

Figure 26: PageRank centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest indegree. The values are calculated using a damping factor of $= 0.85$.

The PageRank centrality is expected to decrease because the total number of nodes N increases. This is a consequence of the initial centrality for the nodes being reduced following the definition of PageRank centrality given by eq. 10. Figure 26 shows that the for every node the index value decreases. The magnitude at which the index value decrease vary for the nodes in the range of $0.04 \quad 0.1$. The nodes that have the highest degree also seem to have their index value decreased the most. There is a correlation between the index value before merge and the magnitude at which the value decreases. The ranking of the index nodes is the same before and after merge, with the exception of 'BIBO stability'. The node's index value decreases more than nodes of similar score, this is due to node not being connected to the super node before merge. Nodes that have the same indegree have similar scores, the difference comes from the score of the connected nodes. The results indicate that every subjects is less influential than they were before, but the ranking of which subject are the most influential is the same, with the exception of 'BIBO stability'.

4.2.3 Katz Centrality

Figure 27: Katz centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest indegree. The values are calculated using a damping factor of $= 0{:}1$ and an initial centrality of $= 1{:}0$.

From Figure 27 it is observed that for each node the index value is reduced after merge. The index value decreases with a magnitude of $0{:}08 \quad 0{:}09$ for all nodes. Compared to PageRank centrality, the magnitude range is much smaller ($0{:}01$ compared to $0{:}06$). The ranking of the index nodes is also the same before and after merge, with the exception of 'BIBO stability', 'Cayley-Hamilton theorem' and 'Jordan normal form' which decreases more than nodes of similar score. As with PageRank centrality, 'BIBO stability' decreases more because it is not connected to the super node before merge. The index value of 'Cayley-Hamilton theorem' and 'Jordan normal form' decreases more than 'System definition' and 'Signal definition' because the latter two nodes have their indegree increase by two, due to the 'Algebra' super node being replaced with two atomic nodes. The results indicate that every subjects is less influential than they were before, but the ranking of which subject are the most influential are the same, with the exception of 'BIBO stability', 'Cayley-Hamilton theorem' and 'Jordan normal form'.

4.2.4 Closeness Centrality

Figure 28: Closeness centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest indegree.

The closeness centrality index decreases for most nodes with the exception of three nodes. The nodes with the highest score are nodes that are connected to the super node. The nodes 'System definition' and 'Signal definition' have their index value increased and the only incoming links are to the super node. They will have their values increased because more nodes can reach them after the super node has been replaced with atomic ones. This is not the case for the other nodes connected to the super node because they have another incoming link from a node. The total shortest path distance, $\sum_{v=1}^{N-1} d(v,u)$, for these nodes is therefore increased and the index value reduces. For the node 'Commutative' the index value is zero before and after merging. This is a consequence of the node having no incoming links. For the atomic nodes that replace the 'Algebra' super node, the index value decreases because the total number of nodes in the graph N increases after the merge, while the number of nodes that can reach these nodes remain the same. The result of the merge is that all subjects are less close to other subjects with the exception of 'Signal definition' and 'System definition' which are a bit closer.

4.3 Connect to nodes with no outgoing edges

This method looks at which atomic nodes has no outgoing edges or an out-degree of zero. This can be interpreted as the ultimate learning goal of the course/subject because it is not needed to learn anything else in the same course. This method is expected to reduce the number of translated atomic edges compared to the naive merge method described in Section 4.1.

Figure 29: The merged graph for the Systems Theory course using the method that connect to nodes with no outgoing edges.

Figure 29 show that each super edge is translated into six atomic edges connected to the nodes 'Jordan normal form', 'Addition', 'QR decomposition', 'LU decomposition', 'Multiplication' and 'Cayley-Hamilton theorem'. Each of these nodes have no outgoing links which can be seen in Figure 7.

4.3.1 Degree Centrality

Figure 30: Degree centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to nodes with no outgoing edges.

The super node is replaced with one or more atomic nodes and hence the degree centrality is expected to increase or stay the same for all nodes. Figure 30 show that the index increases for every node from the Systems Theory course except for 'BIBO stability', which is the result of it being the only node not connected to the super node. The index value also stay the same for the nodes `Matrix multiplication', `Matrix de nition', `Matrix addition' and `Commutative' which are the atomic nodes which does not connect to the nodes in the Systems Theory course. All nodes from the Systems Theory except for `BIBO stability' course have their index value increased by 5, because each super edge is translated into six atomic edges connected to six atomic node from the Algebra course. These six atomic nodes have their index value increased by 6 which equals the degree, deg(Algebra), of the super node. A high amount of atomic edges replace the super edge result in the degree increasing by a high amount for several nodes.

4.3.2 PageRank Centrality

Figure 31: PageRank centrality calculated on the `Systems Theory' learning graph before and after replacing super nodes, using the method that connect to nodes with no outgoing edges. The values are calculated using a damping factor of = 0:85.

The PageRank centrality decreases for all nodes as shown in Figure 31. The behaviour is the same as it was for the method in section 4.2 even though the number of atomic nodes replacing the super node increases. The total number of nodes N in the graph increases after the merge which causes the initial centrality of each node to be lowered, see equation 10. Because the initial centrality for each node is lowered the index value is decreased for the nodes. The nodes `Jordan normal form' and `Cayley-Hamilton theorem' have a higher index value than the other atomic nodes from the Algebra course, because they can be reached by a higher amount of nodes. For this method the ranking of the node centrality index does not remain the same. After the merge, every node from the Systems Theory course have a higher index value compared to the nodes from the Algebra course, with the exception of `BIBO stability'. Both the subjects originally within the course and the subjects merged seem to loose their in uence after the merge. The subjects in the graph before the merge loose less in uence than the subjects from the Algebra course, with the exception of `BIBO stability'. This is a consequence of the subjects being highly in uential within the Algebra course shown in gure 7 and the fact that no new subjects can reach them due to the direction of the super edges.

4.3.3 Katz Centrality

Figure 32: Katz centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to nodes with no outgoing edges. The values are calculated using a damping factor of = 0 :1 and an initial centrality of = 1 :0.

The Katz centrality decreases is expected to decrease for all nodes as it did with Pagerank centrality. Figure 32 show that this is the case and what di ers is the magnitude at which the index change. The six atomic nodes from the Algebra course that have no outgoing edges have the index value decreased by the same amount. The nodes `Cayley-Hamilton theorem' and `Jordan normal form' have a higher index value than the other atomic nodes from the Algebra course because they have have two incoming edges. For the nodes present in the graph before the merge, the index value reduces the least. The only exception is 'BIBO stabilty' which index value is reduced by about 0:14, compared to 0:05. The reason for this is that the node is not connected to a super node before the merge and will have the same neighbors after the merge. The subjects present in the graph before the merge have a little less in uence in the course than before. The subjects that replace the super node also have less in uence in the Systems Theory course than they did in the Algebra course.

4.3.4 Closeness Centrality

Figure 33: Closeness centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to nodes with no outgoing edges.

Figure 33 show that the nodes present in the graph before merge has their index value increased, while the atomic nodes from the Algebra has their index value decreased. The node `Commutative' has an index value of zero because it only has outgoing edges. The nodes present in the graph before merge will have their index values increased because the total number of nodes that can reach it N increases, several nodes from the Algebra course can reach every node from the Systems Theory course and the average shortest path to any other node $\frac{N-1}{\sum_{v=1}^{N-1} d(v;u)}$ increases. The index value for the nodes `System de nition' and `Signal de nition' increases more because the nodes only have an incoming edge from the super node. The index value of `BIBO stability' increases the least because there is two edges between the node and the super node. For the subjects with an index value of zero, it is impossible to tell how close the subject is to other subjects. The subjects are however not dependant on any other subject as they have an in-degree of zero. The subjects present in the graph before the merge are much more central than they were before. The atomic nodes from the Algebra course are however not as close as they were before, although by a small margin.

4.4 Connect to nodes with highest degree

This method looks at which atomic nodes have the highest degree. The degree of a node is given by equation 5 and is equal to the total number of edges or the weighted sum if the edges are weighted. A practical interpretation is to find the subjects which are the most influential in the graph. The number of atomic nodes that have the highest degree is expected to be low.

Figure 34: The merged graph for the Systems Theory course using the method that connect to the nodes with the highest degree.

Figure 34 show that the super edge was replaced with atomic edges connected to only one atomic node. From figure 7 it is observed that the 'Matrix multiplication' node has a degree of $deg(v) = 5$ which is the highest in the graph.

4.4.1 Degree Centrality

Figure 35: Degree centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest degree.

The method results in the super edges being translated into atomic edges connected to only one atomic node from the Algebra course. The degree is expected to only increase for that single node from the Algebra course, which is the case as shown in figure 35. The index value increases by a magnitude of six, because this is the degree of the super node $deg(Algebra) = 6$. The centrality index is unaffected for every other node from the Systems Theory and Algebra course.

4.4.2 PageRank Centrality

Figure 36: Pagerank centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest degree. The values are calculated using a damping factor of = 0:85.

The Pagerank centrality is observed to decrease for every node as shown in gure 36. The magnitude in the decrease vary from the di erent nodes. Nodes from the Systems Theory course decreases more than the nodes from the Algebra course, which is re ected in the ranking of the index values. The reason for this is that there are several nodes that have a high score that links to other nodes. Nodes with a high score that are linked include `Commutative', `Matrix de nition', `Matrix addition' and `Matrix multiplication'.
The node `Periodicity' from the Systems Theory course have a high score because it has a link to the `Signal de nition' node, in addition to the super node. Every subjects loose in uence after the merge, but subjects from the Systems Theory course loose more than subjects from the Algebra course. There is also not a clear distinction of which course has the most in uential nodes, as both courses have subjects with high in uence and subjects with low in uence.

4.4.3 Katz Centrality

Figure 37: Katz centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest degree. The values are calculated using a damping factor of $= 0.1$ and an initial centrality of $= 1.0$.

As with Pagerank centrality, the Katz centrality decreases for every node. The index value for the nodes decreases by either 0.08 or 0.11, depending on what graph the node is from. The nodes from the Systems Theory course decrease by a magnitude of 0.11, because they are connected to the super node. Although `BIBO stability' is not connected to the super node it will still loose centrality because `System de nition' loose centrality. Every node from the Algebra course decrease with a similar magnitude. The nodes that have the highest centrality score have an incoming edge from the `Matrix multiplication' and one other node. After the merge there are three groups of nodes with similar index scores. Subjects from either course become less in uential after the merge, and the index score ranking of the nodes remain the same after the merge.

4.4.4 Closeness Centrality

Figure 38: Closeness centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with the highest degree.

For this method, the closeness centrality index decreases for every node. The index value decreases for the nodes from the Systems Theory graph because the super edge is only replaced with six atomic edges connected to one atomic node from the Algebra course. As a consequence, the nodes from Algebra that can reach the nodes in the Systems Theory graph are `Matrix multiplication', `Matrix de nition' and `Commutative'. Therefore, both the total number of nodes N increases and the average shortest path will decrease, which results in the index value decreasing. The index value for the nodes from the Algebra course will decrease because no new nodes can reach these nodes and the total number of nodes in the graph N increases. The node `Commutative' have an index value of zero both before and after merge because it cannot be reached. All subjects are less close than they were before, except for `Commutative'. The subjects in the Systems Theory course are less close because the nodes in Algebra that can reach them are not close. The subjects in Algebra are equally close to the other subjects in Algebra, but no nodes in the Systems Theory course can reach them and the overall closeness will decrease.

4.5 Connect to nodes with incoming and outgoing edges

This method looks at which atomic nodes have both an incoming edge and an outgoing edge, regardless of either degree. Subjects which have both incoming and outgoing edges are central in the course because they require previous knowledge and they are needed to learn other subjects.

Figure 39: The merged graph for the Systems Theory course using the method that connect to the nodes with incoming and outgoing edges.

Figure 39 show that the super edge is translated into atomic edges connected to the three nodes `Matrix multiplication', `Matrix addition' and `Matrix de nition'. Both the nodes `Matrix addition' and `Matrix multiplication' have an incoming edge from the `Matrix de nition' node which is also present in the graph. Figure 7 show that these are the nodes that have both an incoming and an outgoing edge.

4.5.1 Degree Centrality

Figure 40: Degree centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with incoming and outgoing edges.

The degree centrality is expected to increase because edges are only added. Figure 40 show that the degree increases by a magnitude of two for every node present in the graph, except for `BIBO stability'. The index value for `BIBO stability' does not increase because it is not connected to a super node. The three nodes `Matrix addition', `Matrix multiplication' and `Matrix de nition' from the Algebra course have their index value increased by ve, because $\deg(\text{Algebra}) = 6$. Nodes from the Algebra course that does not have both an incoming and an outgoing edge will have the same index score.

4.5.2 PageRank Centrality

Figure 41: Pagerank centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with incoming and outgoing edges. The values are calculated using a damping factor of $= 0.85$.

For the Pagerank centrality index the values are observed to decrease for each node as shown in figure 41. Pagerank centrality assigns some centrality for each node which is why no node has an index value of zero. The initial centrality for each node decreases because the total number of nodes N increases and the damping factor stay the same, see eq. 10. `Periodicity' has a high score because it has an incoming edge from the 'Signal definition' node which has a high score. The node `BIBO stability' decreases more than the other nodes present in the graph before merge because it gains no new edges after the merge. The ranking order of the index value changes. The index value decreases more for nodes from the Algebra course than from the Systems Theory course. The subjects are not as influential as they were before the merge. Subjects from the Systems Theory course are more influential than the subjects from the Algebra course, with the exception of `BIBO stability'. The subjects from the Algebra course are also less influential in the Systems Theory course than they were in the Algebra course.

4.5.3 Katz Centrality

Figure 42: Katz centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with incoming and outgoing edges. The values are calculated using a damping factor of $= 0.1$ and an initial centrality of $= 1.0$.

The Katz centrality index values are also observed to decrease as with Pagerank centrality. Both Pagerank and Katz centrality are variations of Eigenvector centrality but are calculated differently. The index value for each node decreases with a similar magnitude of 0.08 0.09. The node `BIBO stability' decreases a bit more because it does not have a direct edge to the super node and the connection is therefore weighted less than for the other nodes in the Systems Theory graph. The nodes `System Definition' and `Signal definition' decreases less than the other nodes. As with Pagerank centrality, the ranking order of the index centrality changes. Subjects from the Algebra course are not as influential as subjects from the Systems Theory course, although every subject has less influence in the Systems Theory course after the merge. The node `BIBO stability', which has no direct connection to the subjects replacing the super node, loose more influence than other subjects that have a direct connection.

4.5.4 Closeness Centrality

Figure 43: Closeness centrality calculated on the 'Systems Theory' learning graph before and after replacing super nodes, using the method that connect to the nodes with incoming and outgoing edges.

In contrast to the other indexes the Closeness centrality index values increase for some nodes and decrease for others. Figure 43 show that the index value for nodes present in the graph before merge decreases by a magnitude of $0{:}2$ $0{:}25$, except for the nodes `System de nition' and `Signal de nition' for which the index increases by $0{:}6$. The index value decreases for ever atomic node in the Algebra course, but with di erent magnitude. It decreases the most for the nodes `Cayley-Hamilton theorem' and `Jordan normal form' by a magnitude of $0{:}12$, compared to $0{:}5$ $0{:}7$ for the other nodes. The index value for the `Commutative' node is zero before and after merge, since the node has no incoming edges. The index value increases for `Signal de nition' and `System de nition' because the number for reachable nodes increases, even tho the average shortest path increases. For the other nodes in the Systems Theory course the value decreases for the same reason the other nodes index value increases, but the nodes can be reached by two nodes before merge with a single edge compared to just one node. The index value decreases for every node in the Algebra course because the total number of edges N increase, while the other factors remain the same. For the subjects in the Systems Theory course, some subjects are less close than they were before and others are closer than before. The average combined closeness for the nodes is almost the same. The subjects in the Algebra course are less close than they were before, due to no new subjects being able to reach these subjects.

5 Discussion

This section discusses the results and observations from Section 4. The change in each centrality index is discussed for very method. The resulting graphs for each method is compared to see how well the translated subjects re ect the course dependencies.

5.1 Degree Centrality index performance

(a) Naive method.

(b) Connect to the nodes with the highest indegree

(c) Connect to nodes with no outgoing edges.

(d) Connect to the nodes with the highest degree.

(e) Connect to the nodes with incoming and outgoing edges.

Figure 44: Degree centrality index comparison showing how the index changes for every method.

For each method the degree centrality for each node either increases or stay the same. The only edges that are removed are the super edge connected to a super node, but they are replaced with one or more atomic edge. As a consequence, the methods that replace the super edge with atomic edges which connect to many atomic nodes, will a ect the degree centrality the most. The method that a ect the degree centrality the most is the naive method, for which every index value increases except for `BIBO stability'. This is because the node is the only node not connected to the `Algebra' super node in the Systems Theory learning graph. The centrality metric also increases for most nodes when connecting the super edge with atomic edges to the nodes with no outgoing edges. This method translates the super edge into atomic edges connected to six atomic nodes, for these speci c graphs. The method for which the index increases the least is the method that connect to nodes with the highest degree. There is only one node with the highest degree, so the degree will remain the same for every node from both the Algebra and the Systems Theory course, except for this node. For the last two methods, connecting to the nodes with the highest indegree and connecting to the nodes with both incoming and outgoing edges, the degree centrality increases for half the nodes. For most nodes the index value stay the same or increases by one. The methods connect two and three nodes respectively to atomic nodes in the Systems Theory course, for all of which the degree increases by six.

5.2 Pagerank Centrality index performance

(a) Naive method.

(b) Connect to the nodes with the highest indegree

(c) Connect to nodes with no outgoing edges.

(d) Connect to the nodes with the highest degree.

(e) Connect to the nodes with incoming and outgoing edges.

Figure 45: Pagerank centrality index comparison showing how the index changes for every method.

The Pagerank centrality index value decreases for all nodes and all methods. What differs between the methods, is for which nodes the value decreases the most and how the index ranking of the nodes change. The results from the naive method and the method that connect to nodes with no outgoing edges are similar. The index ranking of the nodes change, with the index value of the nodes from the Algebra course decreasing more than the nodes from the Systems Theory course. The exception for both cases is for 'BIBO stability', for which the index value decreases similarly to nodes from the Algebra course. The index ranking also changes for the method that connect to nodes with incoming and outgoing edges, but with a smaller margin than the other two methods. For the method that connect to nodes with the highest degree, the index rankings barely change. However, not every node from the Systems Theory have a higher score compared to the nodes from the Algebra course as with the two first methods. For the last method, the method that connects to the nodes with the highest indegree the ranking of the nodes stay the same, except for 'BIBO stability'. For every method the lowest scoring node is the 'Commutative' node.

5.3 Katz Centrality index performance

(a) Naive method.

(b) Connect to the nodes with the highest indegree

(c) Connect to nodes with no outgoing edges.

(d) Connect to the nodes with the highest degree.

(e) Connect to the nodes with incoming and outgoing edges.

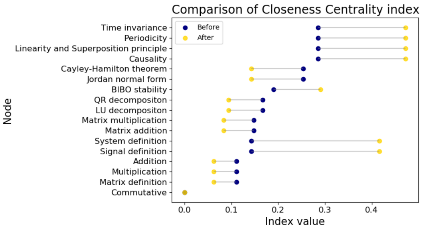
Figure 46: Katz centrality index comparison showing how the index changes for every method.

As with the Pagerank centrality, the index value decreases for all nodes and all methods. What differs between the methods, is for which nodes the value decreases the most and how the index ranking of the nodes change. For the naive method the index value decreases the most for nodes from the Algebra course. The nodes from the Systems Theory course barely decrease in value, except for 'BIBO stability' which decreases in a similar fashion as the subjects from the Algebra course. There is only a small difference between the nodes that are high scoring, as well as between the nodes that are low scoring. The same result is observed with the method that connect to nodes with no outgoing edges. The index value for the nodes from the Systems Theory decreases with a magnitude twice as large as for the naive method. The index value decreases by a similar amount for the nodes from the Algebra course and for the 'BIBO stability' node. The results are similar for the method that connect to nodes with the highest degree and for the method that connect to the nodes with incoming and outgoing edges. The index value decreases by a similar amount for all nodes with the exception of 'BIBO stability', 'Cayley-Hamilton theorem' and 'Jordan normal form', which decreases a bit more. The ranking of the index value stay the same with the exception of the previous nodes mentioned. For the method that connect to the nodes with the highest degree the index value decrease by the same amount for all nodes in the Systems Theory course and by the same amount for nodes from the Algebra course. The index ranking stay the same after merge but there are two clear groups of nodes with almost equal index value. For every method the lowest scoring node is the 'Commutative' node.

5.4 Closeness Centrality index performance

(a) Naive method.

(b) Connect to the nodes with the highest indegree



(c) Connect to nodes with no outgoing edges.

(d) Connect to the nodes with the highest degree.

(e) Connect to the nodes with incoming and outgoing edges.

Figure 47: Closeness centrality index comparison showing how the index changes for every method.

The closeness centrality increases for some nodes and decreases for other nodes, depending on the method. With the naive method, the index values for the nodes from the Systems Theory course increases by a large amount, except for 'BIBO stability' which increases by a small amount. For the nodes in the Algebra course, the index value decreases for every node by a small amount with the exception of 'Commutative' which has an index score of zero. The same behaviour is observed with the method that connect to nodes with no outgoing edges. The difference is that for this method the index value for nodes from the Systems Theory course increase by a smaller amount. For the method that connect to the nodes with the highest degree, the index value decreases for every node except for 'Commutative' which has an index score of zero. The magnitude at which the index value decreases is not dependant on which graph the node was originally from. The behaviour of the method that connect to the nodes with the highest indegree and the method that connect to nodes with incoming and outgoing edges is mixed. For the nodes from the Systems Theory course, the index value increases for 'System definition' and 'Signal definition' but decreases for every other node. For the first method the index value decreases more for the nodes in the System Theory course than the other nodes increase. The behavior is opposite for the method that connect to nodes with incoming and outgoing edges. The nodes from the Algebra course decrease for both methods with a similar magnitude. For every method the lowest scoring node is the 'Commutative' node which has an index value of zero.

5.5 How well do the translation represent the course dependencies

In this section the translated atomic edges and the nodes they connect to are discussed.

The amount of atomic nodes which the translated atomic edges connect to differs for the methods. The naive method connect the atomic edges to every atomic node within the Algebra course. The result is that the atomic edges are connected to ten atomic nodes. This would probably not be the case for real subject dependencies within a course. Every single subject within a course is most likely not a prerequisite for every subject the super node is connected to. On the other hand, the method that connect to the nodes with the highest degree only connect to a single atomic node. This result in the degree being the same for every node except that atomic node. This is also not a likely scenario, all subjects from the Systems Theory course that connect to the super node would most likely not depend on the same single subject. For the method that connect to nodes with highest indegree, the method that connect to nodes with no outgoing edges and the method that connect to nodes with incoming and outgoing edges the number of atomic nodes connected are two, six and three respectively. This is a more likely scenario regarding the course dependencies. The method that connect to nodes with highest indegree, connects to the subjects which require the most previous knowledge. This would be subjects that are complex and depending on complexity of subjects in the Systems Theory course, could provide intuitive connections. The method that connect to nodes with no outgoing edges, looks at which atomic nodes has an outdegree of zero. This would be subjects that are late in the respective course and can be interpreted as the learning outcome of the course, which is an intuitive way to connect the subjects. The last method which connect to the nodes with incoming and outgoing edges, the method connect to subjects that are central in a course. The atomic nodes that have incoming and outgoing edges are subjects that require previous knowledge but is also needed in order to learn something else. From an intuitive perspective, it is a better choice than the naive method and the method that connect to the highest degree, but a worse choice than the other two methods.

5.6 Which method oﬀers the best translation

Each method has benefits and drawbacks, depending on the desired behaviour. The translated atomic edges should result in course and subject dependencies that intuitively makes sense. The centrality metrics also change differently for each method. The naive merge is a poor choice as it translates the super edges into too many atomic edges. The influence of each subject decreases but the method result in subjects from the System Theory course having higher influence than the subjects from the Algebra course. The subjects from the Systems Theory course are also much closer to the other subjects than before, while the subjects from the Algebra course are less close. The method that connects to the highest degree is also a poor choice, because it translates every atomic edge to just a single atomic edge. The influence of each node also decrease and the ranking of the nodes remain almost the same. Subjects from the Systems Theory course are not necessarily more influential than subjects from the Algebra course. All the subjects are also less close to the other subjects than they were before. The method that connect to nodes with the highest indegree make intuitively more sense as it connects to the subjects that are the most complex. The degree increases by a small amount for a few nodes, stay the same for most nodes and increases by a good margin for the nodes with the highest indegree. All subjects from both courses become less influential, by a similar amount. The ranking of the most influential nodes also stay the same. Most subjects are less close to the other subjects. The method that connect to nodes with no outgoing edges is the method that intuitively makes the most sense, as it connects to subjects which can be interpreted as the learning goals of the course. The method does however increase the amount of connection by a large amount. Subjects from both courses become less influential but the subjects from the Systems Theory course are more influential than subjects from the Algebra course. The nodes from the Systems Theory course are much closer to the other subjects than they were before. The subjects from the Algebra course are a little less close to the other subjects. Lastly, the method that connect to the nodes with incoming and outgoing edges is not as intuitive. For a larger course graph the number of nodes that have both an incoming and an outgoing edge is expected to be high. The subjects from both courses are less influential than before but the ranking changes. The subjects from the Systems Theory course are more influential than the subjects from the Algebra course, but by a small margin. Most subjects are less close to the other subjects. Subjects from the Systems Theory course change less than the subjects from the Alegbra course. Depending on the desired effect on the graph, i.e what properties is desired to be preserved, the different methods have positives and negative sides. The choice of method used for translating super edges should be chosen based on the desired effect and interpretation.

5.7 Katz centrality parameterization

It is worth commenting that the result from the Katz centrality is based on the initial centrality and the damping factor . Depending on the desired weighting of the paths, the damping factor can be adjusted.

6 Conclusion

In this thesis we have explored different methods for translating abstract connections in knowledge graphs into more detailed ones. Each method leads to a different result, both in terms of how many connections are added and among which subjects. Summarising, the method that connect to nodes that have the highest indegree resulted in a lowest number of added connections, and to making the influence of the subjects decreasing the least. The method also preserves the ranking of which subjects are the most influential, for almost every subject. This is also observed to happen with the ranking of which subjects are closest to the other subjects. The intuition of the method is that the connections should be based on which nodes require the most previous knowledge.

Relatively to the field example used throughout the thesis, if the desired outcome of the translation is the preservation of the subjects from the Systems Theory course, the method that connect to nodes with no outgoing edges is a valid method. The method results in a lot of connections, which is not an ideal outcome, but the influence and closeness of the subjects from the Systems Theory course are high. The closeness will increase for every subject within the Systems Theory course but decrease with a small margin for subjects from the Algebra course.

Applying the methods to a pedagogical problem such as learning graph for higher education, the methods can aid in translating abstract connections between courses. The methods gives an intuition of how a course is connected in terms of subjects and the relation between them, as well as how subjects from different courses relate. The graph models are subject to personal interpretation, as the CFMs are designed by hand, which should be kept in mind.

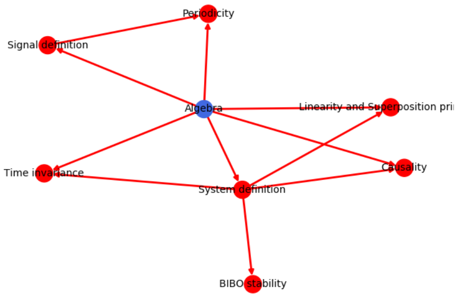
Summarizing, the proposed methods enable to interpret and fuse teachers opinions about the learning outcome into higher detail. Teachers or the programme board of a university can thus use the propose methods to find the appropriate connection between courses and subjects. The results could then be used to help structure semesters and study programmes.

6.1 Future Work

Throughout this thesis the methods are tested on unweighted graphs. The application of weights in a knowledge graph is used to indicate how relevant a subject is to its connected components. An analysis of how well the methods perform for weighted knowledge graphs would provide additional insight into their performance. Implementing methods based on other metrics, such as flow, could provide additional information into which metric the translation should be based upon.

A Figures

A.1 Graph comparison



(a) The original learning graph.

(b) Naive method.

(c) Connect to the nodes with the highest indegree.

(d) Connect to nodes with no outgoing edges.

(e) Connect to the nodes with the highest degree.

(f) Connect to the nodes with incoming and outgoing edges.

Figure 48: The Systems Theory learning graph compared to the resulting graph for each method.

References

[1] Norman Biggs. *Algebraic Graph Theory*. Second. Cambridge: Cambridge University Press, 1993.

[2] Phillip Bonacich. "Simultaneous group and individual centralities." eng. In: *Social networks* 13.2 (1991), pp. 155–168.

[3] Ulrik Brandes. "A faster algorithm for betweenness centrality." eng. In: *The Journal of mathematical sociology* 25.2 (2001), pp. 163–177.

[4] Sergey Brin and Lawrence Page. "The anatomy of a large-scale hypertextual Web search engine." eng. In: *Computer networks and ISDN systems* 30.1 (1998), pp. 107–117.

[5] Reinhardt Diestel. *Graph Theory, Electronic Edition*. New York: Springer-Verlag Heidelberg, 2005.

[6] Linton C Freeman. "Centrality in social networks conceptual clarification." eng. In: *Social networks* 1.3 (1978), pp. 215–239.

[7] LINTON C FREEMAN. "Centrality in social networks I. Conceptual clarification." eng. In: *Social networks* 1.3 (1979), p. 215.

[8] Martin Grandjean. "A social network analysis of Twitter: Mapping the digital humanities community." eng. In: *Cogent arts humanities* 3.1 (2016), p. 1171458.

[9] Steven A Greenberg. "How citation distortions create unfounded authority: analysis of a citation network." eng. In: *BMJ* 339.7714 (2009), S659–213.

[10] Leo Katz. "A new status index derived from sociometric analysis." eng. In: *Psychometrika* 18.1 (1953), pp. 39–43.

[11] Steffi Knorn et al. "Quantitative analysis of curricula coherence using directed graphs." eng. In: *IFAC-PapersOnLine* 52.9 (2019), pp. 318–323.

[12] Kenneth R Koedinger, Albert T Corbett, and Charles Perfetti. "The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning." eng. In: *Cognitive science* 36.5 (2012), pp. 757–798.

[13] M. E. J. Newman. "Mathematics of Networks." In: *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan UK, 2016, pp. 1–8. URL: `https://doi.org/10.1057/978-1-349-95121-5_2565-1`.

[14] Damiano Varagnolo et al. "Graph-theoretic approaches and tools for quantitatively assessing curricula coherence." eng. In: *European journal of engineering education* 46.3 (2021), pp. 344–363.

[15] Emil Wengle. *Modelling Hierarchical Structures in Networks Using Graph Theory : With Application to Knowledge Networks in Graph Curricula*. eng. 2020.

[16] Emil Wengle, Steffi Knorn, and Damiano Varagnolo. "COnCUR - COherence in CURricula: A tool to assess, analyze and visualize coherence in higher education curricula." eng. In: *IFAC PapersOnLine* 53.2 (2020), pp. 17598–17603.