VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

CONCEPTUAL NETWORK OF TEACHERS' TALK: AUTOMATIC ANALYSIS AND QUANTITATIVE MEASURES

Daniela Caballero[1], Toni Pikkarainen[2], Roberto Araya[1], Jouni Viiri[2], Catalina Espinoza[1]

[1]Center for Advanced Research in Education, Institute of Education,

Universidad de Chile

[2]University of Jyväskylä

ABSTRACT

Educational field can take advantage of the improvements of Automatic Speech Recognition (ASR), since we can apply ASR algorithms in non-ideal conditions such as real classrooms. In the context of QuIP project, we used ASR systems to translate audio from teachers' talk into text to study conceptual networks based on what the teacher says during his/her lecture, particularly the key concepts mentioned and their temporal co-occurrence. In the present study, quantitative metrics are provided, such as centrality measures and PageRank, which can be used to analyse the conceptual networks in a broaden way. With a case-study design, two teachers' talk are described quantitatively and qualitatively using the metrics, suggesting that PageRank could be a good metric to find differences in teachers' talk. Finally, we discuss about the potential of this kind of analysis.

INTRODUCTION

The connection of instructional content with students learning has been an active research area for decades. Already in 1970s researchers found that students learning gain was related to the structure of instructional material (e.g. Shavelson, 1972; Geeslin & Shavelson, 1975). In physics education research, Müller and Duit (2004) noticed that the amount of connections between the content structure elements (e.g. definitions, examples, applications, experiments) correlated positively with students' learning gains. Also in a video based study in mathematics teaching, Klieme et al. (2009) found a connection between content aspects and students' learning. Drollinger-Vetter and Lipowsky (2006) showed that both the occurrence of key concepts and the quality of content correlated significantly with students' learning gains (Klieme et al., 2009).

The described analysis of content structure gives rich information of a part of a lesson, but it is time consuming if done manually. The time increases dramatically if several teachers are studied in detail. We can take the advantage of technology to help both researchers and teachers to have a quick feedback about a particular teacher by looking how the concepts are presented and how they are related. A previous work from Helaakoski and Viiri (2014) focused on developing a system that can automatize the transcription of a teachers' talk from the audio of the lesson and the summarization or feedback of how the contents were presented by the teacher.

The research aim of this study is to describe, automatically, the content and content structure of teacher's talk trough content network analysis in a global (and quantitative) and in a more detailed (and qualitative) way. In particular, we have the following Research Question (RQ): How can we describe a lesson, from the teacher's talk, in terms of number of concepts mentioned and their relationship?

To answer our RQ, we studied different network measures using Social Network Analysis such as number of nodes, number of edges, density, diameter, average clustering, average degree and average degree centrality. Also, we used PageRank, a well-known algorithm to search web-pages to describe qualitatively the concept network. Finally, we contrasted two lessons as case-study using the SNA metrics to see whether differences in networks measures and PageRank can give a clue about different talking patterns from the teachers.

The paper is structured as following: in the Theoretical framework we discuss the scope and results from a previous project named QuIP and give a detailed explanation about network analysis with different quantitative measures. Later, we show the methodology of our case-study and its results. Finally, we discuss the implications of the results and give some final remarks of the research in the Discussion and Conclusion section.

THEORETICAL FRAMEWORK

QuIP Project

In the study QuIP (Fischer, Labudde, Neumann & Viiri, 2014) researchers from German, Finland and Switzerland identified instructional patterns, described differences and similarities between physics instruction in the three countries. The aim was to find if there are any relations between these patterns to students learning gains resulted from the pre- and post-test comparisons. One specific aim of the QuIP project was to analyse the content structure of videotaped lessons and to relate different aspects of content structure to students' learning gains. From teachers' speech researchers constructed manually networks showing the connections between concepts. Two concepts were connected if they co-occurred in a 10-seconds time window. Helaakoski and Viiri (2014) found that the number of different physics concepts and connected concept pairs had the highest correlations with the learning gains.

Network analysis

A network is defined as a set of nodes and edges. For the QuIP project and this paper we define a node as a concept mentioned by the teacher. If two concepts are mentioned in the same 10-seconds time window, they will be connected by an edge. Thus, the nodes are concepts and edges are temporal co-occurrence between two pair of nodes.

Many network measures take into account the network structure. The simplest ones are the statistical measures, like the number of nodes and edges. In a previous work from Vargas et al. (2018), the authors used Social Network Analysis (SNA) metrics to show that students' academic performance is correlated with centrality measures of a collaboration network. As Vargas et. al (2018), in this paper we will do SNA on the concept network with the following list of network measures:

- Number of nodes: how many different concepts the teacher said and were related with other concepts
- Number of edges: how many different concepts were said in the same time window
- Density: proportion between number of connections and number of all possible connections. A network with no edges has a density of 0 and a complete network has a density of 1
- Diameter: the greatest distance between any pair of concepts of the network. The distance of two nodes *a* and *b* is the number of nodes that has to be visited from the node *a* to the node *b*.
- Average clustering: the average tendency of the concepts' neighbours to cluster together. It is calculated by dividing the number of connections between the neighbours by the total possible connections between the neighbours
- Average degree: the average of how many linked concepts has each concept of the network
- Average degree centrality: The average of the proportion of how many neighbours does a concept have, normalized by dividing by the maximum possible degree in a simple graph of n-1 nodes where n is the number of nodes in the original network

PageRank

The measures above help us to describe the structure of the network. For example, we can say how many different concepts did the teacher mention, how many different connections were or how far are two different concepts. These measures treat all nodes and edges as equally important, and thus, provide a general description of the network. But they do not indicate which region of the whole network is worth to look in more detail.

In order to find which nodes should have more attention in the analysis we can take a well-known algorithm in web-pages searching and ranking: PageRank algorithm. PageRank gives a global "importance", ranking the web-pages considering the link structure of the Web and not the content of the web-pages (Page, Brin, Motwani, & Winograd, 1999). The algorithm uses a hypothetical random surfer, which is moving through the Web. The pages where the appearance of the surfer have high likelihood are more important (Gleich, 2015) and the links of the Web work as endorsements of the web-pages (Richardson, Prakash, & Brill, 2006). A web-page is important if it has large amounts of endorsements, and also, a web-page is "important" if another "important" web-page endorsements the first one.

The main application of PageRank is searching, but it has many other applications (Gleich, 2015), like summarization of text documents (Gambhir, & Gupta, 2017). We use the algorithm to rank the concepts mentioned by the teacher. If a concept has a high PageRank, then it is "important" to the whole teacher speech. Compared with the frequency of a concept, PageRank takes into account the connection between concepts. With this, we are able to find which concepts are the "core" of teacher's talk.

METHODOLOGY

Data

Two teachers out of 25 were selected from the QuIP Project data, as a case study. The data was 90 minutes of videotaped class of each of the teachers. Both teachers taught the same content: the introduction to the relation between electrical energy and power in ninth grade with no extra guidelines or materials.

The QuIP project developed an instrument to measure students' change in content knowledge. The main topic of the test was the concepts of electrical energy and power. The test included items of multiple-choice and open-ended questions and calculations. As shown in [name deleted to maintain the integrity of the review process] the instrument used a six-level complexity model and it was validated by experts. Students from selected teachers took the pre- and post-test. The booklets for Pre-test had 18 items whereas the booklets for Post-test had 36 items.

The criterion of selection was the effect size of the students' learning gains, calculated as the Cohen's d. First the percentage of achievement of each test (i.e. pre-test score/18 and post-test score/36) was determined. Later, the means and standard deviation were calculated. Students from Teacher A had an effect size of 1.310, whereas students from Teacher B had an effect size of -0.005. Thus, in terms of their respective students' learning gains, Teacher A was the teacher with greatest effect size and Teacher B had the lowest effect size from the whole QuIP project.

Data analysis

Data analysis consisted several steps in order to create an automatic concept network and its quantitative measures. The input of each step is the output of the previous one, except for the ASR which used the videotaped classes of the teacher as input. The Figure 1 shows the pipeline of Data analysis.

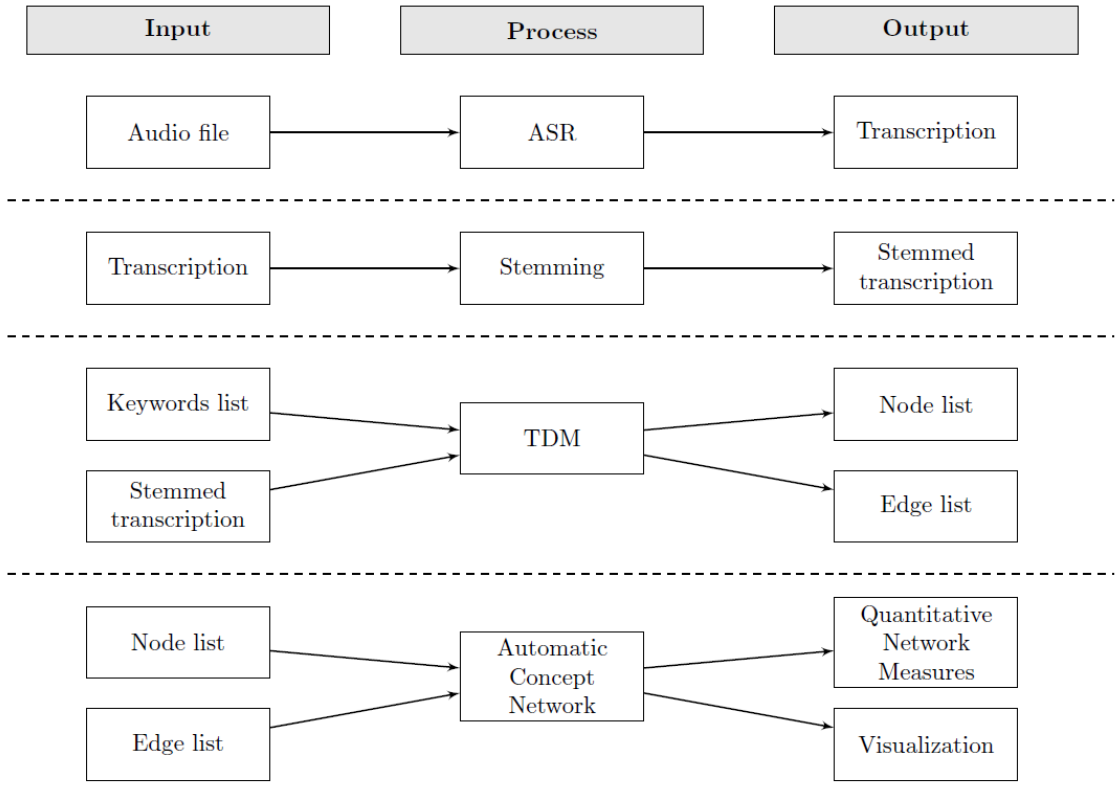


Figure 1. Pipeline of Data Analysis

Automatic Speech Recognition (ASR)

An Automatic Speech Recognition (ASR) algorithm developed by the Aalto University (Kronholm et al., 2017; Caballero et al., 2017) was ran on the two videotaped classes. The output of the process is the transcription of the teacher's speech in a text form. The following lines are an example of ten seconds transcription (each line is a 5 second transcription):

> 2430.0 2435.0 mutta niitä elektroneja silloin se palaa kirkkaammin ihan oikeilla linjoilla oli
>
> 2435.0 2440.0 vaikuttaako jännitettä on pyörimisen teho

It is clear that the transcription is not 100% accurate. There is no transcription system (automatic or human) that have perfect performance (Blanchard et al, 2015; Hazen, 2006). Since we are only looking at keywords (as shown in the Text Data Mining step) it is not imperative for us that all the words be transcribed

correctly, nor the tense or if it is plural or singular. For the analysis it is important to get the root of the concepts, as explained in the following section.

Text Data Mining (TDM)

The first step of text mining process is pre-processing the input, the method used was stemming all the words in the transcription. The process was automated using Natural Language Toolkit for Python (NLTK), particularly the Snowball Stemmer which uses the algorithms developed by Porter (1980). In the stemming method, the root of a word is identified so the number of words is reduced (Vijayarani, Ilamathi, & Nithya, 2015). For example, the words *jännitteellä*, *jännitteen* and *jännite* can be stemmed to the word *"jännit"*. In the example of the transcription shown in the previous section, the stemmed concepts are "elektron" (from the word *elektroneja*, electrons), "jännit" (from the word *jännitettä*, voltage) and "teho" (from the word *teho*, power) was found in the teacher's speech.

The stemmed text form of the speech was analysed by a text mining process. The text mining process was used to later analyse the words used in the teachers' talk. To narrow the set of words to be looked up, one of the researchers, who is also a physics teacher, listed 486 different physics concepts from textbooks. This process consisted in taking the concepts from a concept directory of a whole high school physics book. All the concepts were considered as keywords and the list was the input of the text mining process. The process looked for keywords and their connections in a 10 seconds window time. A connection between two words means that the words were said in the same 10 seconds window time.

Automatic Concept Network

The automatic concept network was done using the keywords as nodes and keywords connections as edges found in the text. The process was done with a Python script using the `networkx` module, which handles the visualization and metrics calculation with nodes and edges as input. Figure 2 shows an example of a mock-up concept network with five concepts A, B, C, D and E.

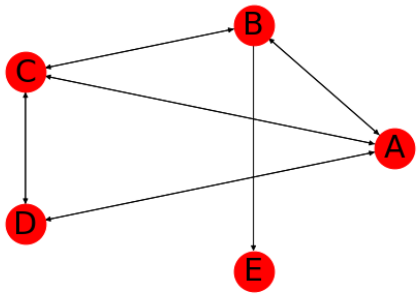


Figure 2. Simple example.

Quantitative Network Measures

Network measures shown in the Theoretical framework section were calculated: number of nodes, number of edges, average clustering, average degree centrality,

density, diameter and average degree. Using the network example shown in Figure 1, Table 1 shows the network measures.

Table 1. Network measures for Network shown in Figure 1.

Network Measure	Example
Number of nodes	5
Number of edges	11
Average degree centrality	1.1
Average degree	4.4
Diameter	3
Clustering	0.53
Density	0.55

Finally, the PageRank algorithm was used to find the most important keywords said by each of the teachers. In the example network of Figure 1, the PageRank is 0.256 for concept A, 0.193 for concept B, 0.256 for concept C, 0.193 for concept D and 0.102 for concept E. This means that the most important concepts are A and C (equally important), followed by the concepts B and D (equally important). The least important word is E.

Figure 3 shows an example of the process, using the transcription of the ASR subsection.

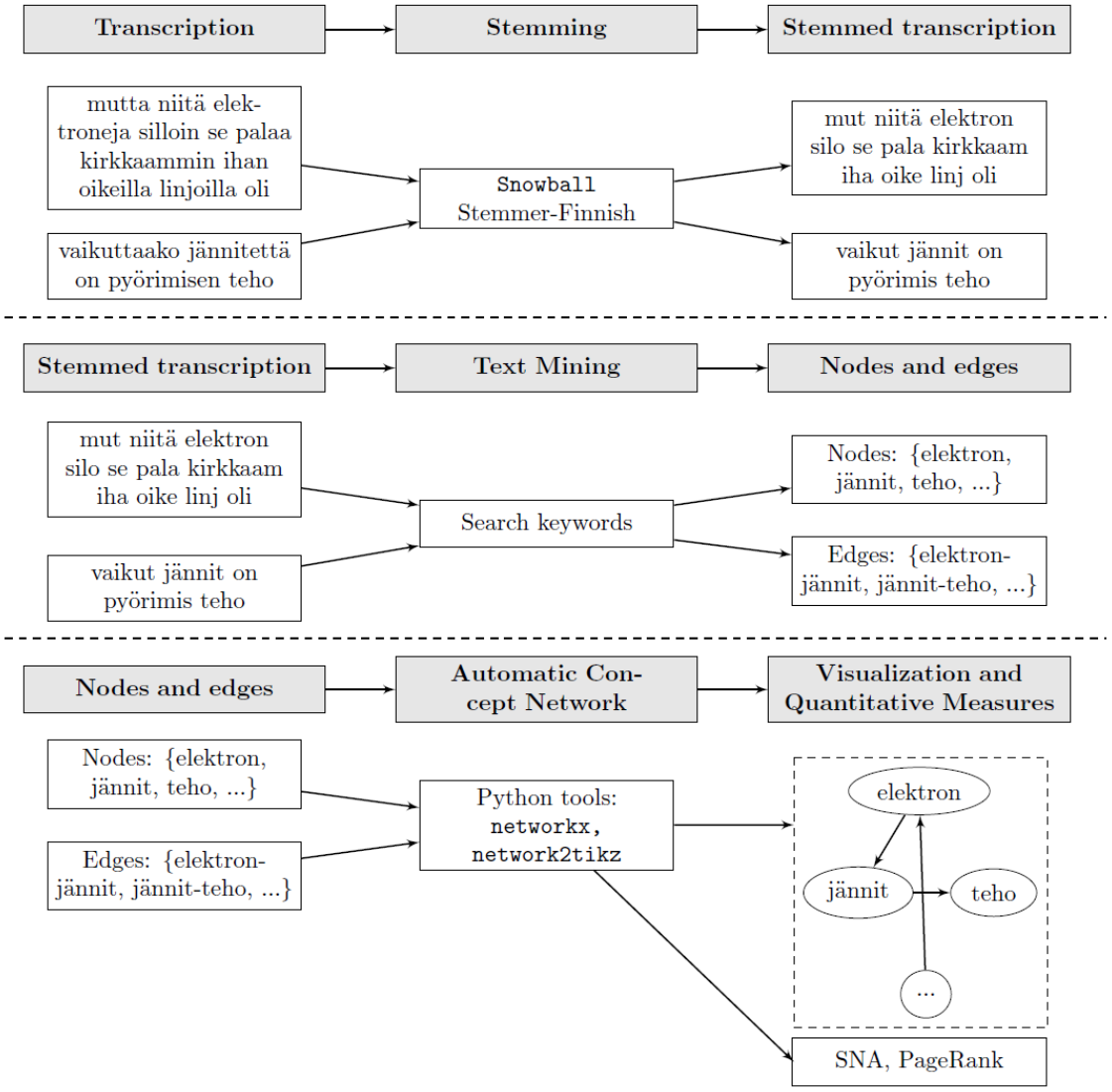
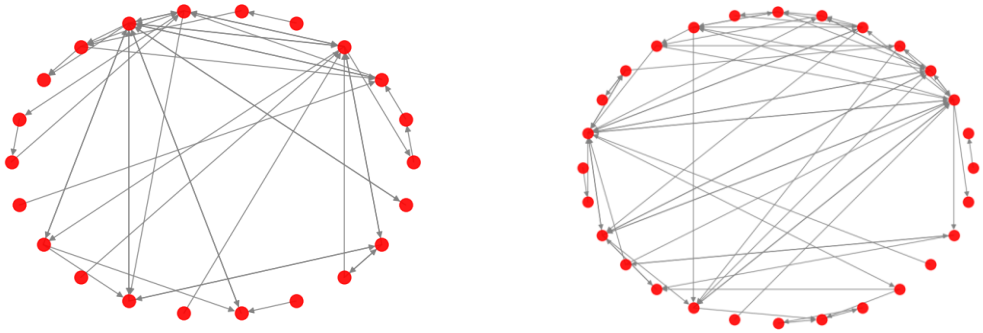


Figure 3. Data analysis with an example

RESULTS

Figure 4 depicts the automatic concept networks for Teacher A and Teacher B. Each node (red dot) is a concept and the edges (arrows) is a connection of two pair of concepts. For the simplicity of the image, the nodes are not shown with labels, but the networks that teachers can see have the concept label. The network is a directed graph, where the direction of the edge gives temporal relationship between two concepts. The network measures calculated for both networks are shown in Table 2.



(a) (b)

Figure 4. (a) Teacher A and (b) Teacher B automatic concept network.

Table 2. Network measures for Teacher A and B.

Measure	Teacher A	Teacher B
Number of nodes	22	28
Number of edges	45	69
Average degree centrality	0.195	0.183
Average degree	4.091	4.929
Diameter	5	5
Clustering	0.236	0.199
Density	0.097	0.091

From the data shown in Table 2, the classes of both teachers can be described as classes with a lot of concepts mentioned (also similar number of nodes). Teacher B connected more different concepts than Teacher A (69 edges versus 45). Also, for Teacher B each of the concepts were connected, in average, to 5 different concepts (4.929), whereas Teacher A each of the concepts were connected, in average, to 4 different concepts (4.091). All the other measures are quite similar.

The top 5 PageRank for the Teachers A and B are shown in Table 3. Data in Table 3 shows the Teacher A was more "theoretical", whereas Teacher B was more "concrete". The most important word in both cases was the stemmed version of lamp, but Teacher A connected it with keywords considered as more theoretical and Teacher B connected it with keywords more concrete or practical. The keywords in Teacher A PageRank list are also the main physics concepts to teach the power of electricity since Power = Voltage x Current.

Table 3. The top 5 PageRank concepts for the Teachers A and B

Rank	Teacher A		Teacher B	
	Word	Value	Word	Value
1	Lamp	0.235	Lamp	0.123
2	Voltage	0.151	Battery	0.111
3	Power	0.136	Voltage	0.096
4	Current	0.075	Series	0.092
5	Circuit	0.051	Short circuit	0.068

As seen also in Table 3, there is a 64.8% of chances that teacher A is talking about these 5 concepts. For teacher B the chances drop to 49.0%. Although both teachers have similar amount of concepts mentioned, Teacher A focused mainly in fewer concepts than Teacher B.

On the other hand, Teacher A made a total of 105 connections, with 45 unique connections (edges in Table 2). Teacher B made a total of 153 connections, with 69 unique connections. Table 4 shows the number of connections for each pair of concepts of the top PageRank concepts; the edge connects a concept that was mentioned at first (from) with another concept said afterwards in the same 10-seconds time window (to).

Table 4. The connection between concepts for the Teachers A and B

	Teacher A			Teacher B		
	From	To	Connections (percentage)	From	To	Connections (percentage)
1	Voltage	Lamp	13 (12.38%)	Battery	Lamp	13 (8.49%)
2	Lamp	Power	9 (8.57%)	Battery	Series	10 (6.53%)
3	Lamp	Voltage	5 (4.76%)	Voltage	Lamp	8 (5.22%)
4	Voltage	Power	5 (4.76%)	Battery	Voltage	6 (3.92%)
5	Power	Voltage	5 (4.76%)	Series	Battery	5 (3.26%)
6	Lamp	Circuit	4 (3.80%)	Series	Lamp	5 (3.26%)
7	Power	Lamp	4 (3.80%)	Lamp	Series	4 (2.61%)
8	Circuit	Voltage	2 (1.90%)	Voltage	Battery	4 (2.61%)
9	Circuit	Power	2 (1.90%)	Lamp	Voltage	3 (1.96%)
10	Lamp	Power	2 (1.90%)	Series	Voltage	2 (1.31%)
11	Power	Voltage	1 (0.95%)	Lamp	Battery	2 (1.31%)
12	Power	Lamp	1 (0.95%)	Voltage	Series	1 (0.65%)

From Table 4 we can see that the most frequent connections made by Teacher A related the three keywords Voltage, Lamp and Power. The Teacher B related the keywords Lamp, Series, Voltage and Battery. As described with the PageRank, Teacher A talked about mainly concrete keywords and connected them with few other keywords. Teacher B talked about mainly practical keywords and connected them with more keywords.

DISCUSSION AND CONCLUSIONS

Content-structure analysis of teachers' speech is important in order to understand its connection/relation to students' learning gains. In previous studies, it has been found that the number of different physics concepts and connected concept pairs had the highest correlations with the learning gains. Doing this analysis manually is a challenging task, because it is time consuming and expensive. The aim of this paper was to find a way to describe a lesson, from the teacher's talk, in terms of number of concepts mentioned and their relationship. We used SNA metrics and PageRank to describe two teachers' talk considering how the concepts were distributed during the session.

With these approaches we describe the teachers' talk as follows: Teacher A used the same amount of keywords as Teacher B and connected quite less words together than Teacher B. PageRank analysis reveals that the Teacher B focused more in concrete concepts, whereas the Teacher A related the concept lamp with other theoretical concepts. Moreover, Teacher A concepted lamp with few other concepts, whereas Teacher B used her/his time to relate the same word lamp with more concepts, and also more practical ones This could explain the differences in the effect size of both teachers. These results say that there is a threshold where is good to have more concepts mentioned, but if those concepts are not connected with more central keywords, then the message received by students could be weaken. Thus, students could not perform better in the post-test.

This is an explorative case study which considered only two teachers. It is a limitation, since we cannot generalize the results, but it is an initial step to explain some differences among Teachers' performance, in an automatic way. This study shows the potential of the methods, but it needs to be proven with a bigger sample size. If developed even in a more automatic and handy way we could have possibilities in real classroom context when helping teachers. There are concrete possibilities in teacher education and teacher evaluation. In particular, this analysis can be helpful to discuss how the teacher is distributing her/his time in the class when lecturing, and whether she/he is presenting the concepts with the strength it is planned, depending on the objectives of the lesson and to see what is the scope of the lesson, if more experimental, calculus, theoretical or practical oriented. It also gives the core concepts of the lesson, and it can be a way of summarization a large amount of data without looking in detail the audio or video of the lesson performed by teachers.

There are several lines of future work. First, we have to find ways to compare PageRank values for different teachers. PageRank depends on the size of the network. Thus, we would like to compare the content networks of teachers with large size with teachers with smaller network size. Second, there is still a need to compare the manual qualitative description of lessons with PageRank. Third, if the same analysis is done in a random 5-minutes of the teacher's talk, which results would appear? Finally, it would be interesting to see the dynamics of teacher's talk: how PageRank changes in time for different teachers.

ACKNOWLEDGEMENTS

Support from ANID/PIA/Basal Funds for Centers of Excellence FB0003 and ANID-FONDECYT grant N° 3180590 is gratefully acknowledged. The authors are also thankful for the funding provided to project No. 294218 by the Academy of Finland

REFERENCES

Blanchard, N., Brady, M., Olney A. M., Glaus, M., Sun, X., Nystrand, M., Samei, B., Kelly, S., D'Mello, S. (2015) A Study of Automatic Speech Recognition in Noisy Classroom Environments for Automated Dialog Analysis. In: Conati C., Heffernan N., Mitrovic A., Verdejo M. (eds) Artificial Intelligence in Education. AIED 2015. Lecture Notes in Computer Science, vol 9112. Springer, Cham

Caballero D. et al. (2017) ASR in Classroom Today: Automatic Visualization of Conceptual Network in Science Classrooms. In: Lavoué É., Drachsler H., Verbert K., Broisin J., Pérez-Sanagustín M. (eds) Data Driven Approaches in Digital Education. EC-TEL 2017. Lecture Notes in Computer Science, vol 10474. Springer, Cham

Chen,Y. N., Huang Y., Kong S.-Y., & Lee L. (2010). Automatic key term extraction from spoken course lectures using branching entropy and prosodic/semantic features. *In proc. IEEE Spoken Language Technology Workshop*, pp. 265-270.

Drollinger-Vetter, B. & Lipowsky, F. (2006). Fachdidaktische Qualität der Theoriephasen. In E. Klieme, C. Pauli, & K. Reusser (Eds.) *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"*. Materialien zur Bildungsforschung (Vol. 15). Frankfurt am Main, Germany: GFPF.

Fischer, H.E., Labudde, P., Neumann, K., & Viiri, J. (2014) Quality of instruction in physics. Comparing Finland, Germany and Switzerland. Münster: Waxmann.

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review, 47*(1), 1-66.

Geeslin, W. E. & Shavelson, R. J. (1975). Comparison of content structure and cognitive structure in high school students' learning of probability. *Journal for Research in Mathematics Education,* 6 (2), 109–120.

Gleich, D. F. (2015). PageRank beyond the Web. *SIAM Review*, *57*(3), 321-363.

Hazen, T. J. (2006). Automatic alignment and error correction of human generated transcripts for long speech recordings. In *Ninth International Conference on Spoken Language Processing*.

Helaakoski, J. & Viiri, J. (2014) Content and content structure of physics lessons and students' learning gains. In H.E. Fischer,P. Labudde, K. Neumann, & J. Viiri, (2014) Quality of instruction in physics. Comparing Finland, Germany and Switzerland. Waxmann, Munster. pp. 93-110.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study. Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janík & T. Seidel (Eds.) *The power of video studies in investigating teaching and learning in the classroom* (pp 137– 160). Münster: Waxmann.

Kronholm, H., Caballero, D., Mansikkaniemi, A., Araya, R., Lehesvuori, S., Pertilä, P., Virtanen, T., Kurimo, M., & Viiri, J. (2017). The automatic analysis of classroom talk. FMSERA Journal, 1(1), 142-151.

Muller, C. T. & Duit, R. (2004). Die unterrichtliche Sachstruktur als Indikator fur Lernerfolg – Analyse von Sachstrukturdiagrammen und ihr Bezug zu Leistungsergebnissen im Physikunterricht. *Zeitschrift für Didaktik der Naturwissenschaften,* 10, 146–160.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14* (3), 130-137.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology,* 63 (3), 225–234.

Richardson, M., Prakash, A., & Brill, E. (2006). Beyond PageRank: Machine Learning for Static Ranking

Wald M., & Bain K. (2008). Universal access to communication and learning: the role of automatic speech recognition. *Universal Access in the Information Society*, 6 (4), 435-447.

Wang, Z., Pan, X., Miller, K. F. &. Cortina K. S (2014). Automatic classification of activities in classroom discourse. *Computers & Education* 78, 115-123.

Vargas, D. L., Bridgeman, A. M., Schmidt, P. B., Wilcox, B. R., & Carr, L. D. (2018). Correlation between student collaboration network centrality and academic performance. *Physical Review Physics Education Research, 14*(2), 020112.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks, 5*(1), 7-16.