British Journal of Educational Technology Vol 0 No 0 2020 1–1
doi:10.1111/bjet.12938

The network approach to assess the structure of knowledge: Storage, distribution and retrieval as three measures in analysing concept maps

Anastasiya Kapuza , Ismo T. Koponen and Yulia Tyumeneva

Anastasiya Kapuza is a junior research fellow and PhD student at Institute of Education, National Research University Higher School of Economics. Her research focus is the use of digital technologies in education for learning and assessment. Ismo T. Koponen is a professor of physics, in the Department of Physics, University of Helsinki, Finland. He is teaching conceptual physics courses for pre-service physics teachers, and his research interests are focused on concept learning. In his research he uses and develops network methods and approaches for science education research. Yulia Tyumeneva is a Senior Research Fellow in Institute of Education, National Research University Higher School of Economics. Earned her PhD in Psychology in the Institute of Psychology of Russian Academy of Education. Her research interests are learning and thinking, analogy, concept acquisition and assessment in cognitive area. Address for correspondence: Anastasiya Kapuza, Institute of Education, National Research University Higher School of Economics, 20 Myasnitskaya ulitsa, Moscow 101000 Russia. Email: a.v.kapuza@gmail.com

Abstract

We present three new standardised network concept map (CM) measures that can provide unique information about learning-related progress, which cannot be determined from previously known measures. Grounded in cognitive development theory on the one hand, and network theory on the other hand, our measures reveal how knowledge is stored, distributed and retrieved. We validated the new measures by testing their ability to discriminate between CMs of respondents with different levels of competency in statistics (students before and after taking an introductory statistics course and experts in the field of statistics). We also validated our measures against the most commonly used traditional and network measures. Based on a small sample of respondents, we show that two of the newly proposed compound measures reveal significant differences between experts and novices in the field, with higher values for experts, showing that expert knowledge is better distributed, more connected and balanced. More importantly, our measures were sensitive enough to show learning-related progress for students, albeit statistically non-significant, while common indicators from network theory did not demonstrate these small shifts. The validity of our new measures can be inferred from the consistency of the results from different sets of measures.

Introduction

While the usefulness of concept mapping as an assessment tool is well established, its applicability remains limited. An important reason for this can be the lack of objective indicators for scoring concept maps (CMs). Basically, the number of concepts in a map and the links between the concepts are used only as objective measures. All other traditional indicators, such as accuracy (Ruiz-Primo, Schultz, Li, & Shavelson, 2001), "quality" (Novak & Gowin, 1999), etc., heavily rely on separately scored expert ratings (usually, teachers of these students). Even if criteria and rubrics are provided for evaluating maps, experts inevitably inject an undesirable degree of subjectivity,

Practitioner Notes

What is already known about this topic

- A concept map (CM) is a knowledge assessment tool that can be used to evaluate student understanding of a topic and learning-related progress.
- However, there has been a lack of standardised, complex and domain-general CM metrics.
- Network theory has been shown to be useful in developing a standardised approach to CM evaluation.

What this paper adds

- Based on the theory of cognitive development, we derived three standardised network measures of a CM: knowledge storage capacity, knowledge distributivity and knowledge retrieval.
- The new measures were empirically found to discriminate between groups of students with different levels of competency.

Implications for practice and/or policy

- The new measures provide unique information about learning-related progress.
- The newly developed measures enable educational researchers to make more objective evaluations of individual map elements, while simultaneously allowing for a holistic view of the quality of the knowledge structure of students.

which affects the reliability of the scores (Watson, Pelkey, Noyes, & Rodgers, 2016). In addition, traditional map indicators are mostly subject-specific or even theme-specific. This greatly reduces the generalisability and comparability of findings from different subject domains.

To address these problems, we developed a method of analysis of CMs that allows researchers to reliably detect learning-related changes in student knowledge structure. Based on theories of cognitive development and network theory, we derived three new standardised network CM measures that reflect how knowledge is stored and distributed in a knowledge structure, and how that knowledge is available for retrieval. These three measures are compound in the sense that they are formed based on more elementary network measures, which are directly related to the linkage structure of the network. The elementary measures, however, are not discriminative enough alone to provide a basis for classification of the CMs, whereas the compound measures are. Our goal was to provide structure-based information about learning-related progress, which cannot be obtained from previously known methods. To support the validity of the newly derived measures, we checked whether they discriminate between respondents with different levels of competency in a field, and show results consistent with other measures of CMs.

In the next sections, we describe how theories of cognitive development and learning guide our search for unified CM measures. Then, we explore the potential of network theory as a source for more objective measures of knowledge structure. Furthermore, we derive new compound measures based on the already known network measures.

Restructuring knowledge as a result of learning

We wish to explore what theories of learning and cognitive development can tell us about learning-related changes in knowledge structure, and if it can differentiate between less and more developed structures.

Based on Vygotsky's theory, the development of everyday and scientific concepts are different. Everyday concepts develop from concrete abstract generalised meanings. In contrast, scientific concepts develop from initially abstract meanings (concepts), to the awareness about the concrete objects and phenomena that scientific concepts refer to (Vygotsky, 1982). By this logic, the more developed a knowledge structure is, the more levels of generalisability that are involved. For example, the concept of correlation might initially be represented as only formal statistics unconnected with other concepts and poorly exemplified. After one experiences various situations applying correlation, the concept might become interrelated with other concepts, ranging from the abstract (such as a regression) to the very concrete (such as a weather forecast).

Piaget suggests that new information is assimilated by an existing structure to the extent and in the manner that the structure is able to assimilate itself (Piaget, 1952). At the same time, the information being assimilated modifies the existing structure to the extent that the new information mismatches with the old information. Inevitable mismatches feed the ongoing processes of rebuilding the cognitive structure and forming "layers" of concepts, where new concepts form higher layers, which are built on and absorbed by the concepts from lower layers.

What follows from these fundamental theories is that the availability of concepts from multiple generalisability levels seems to be an important indicator for a more developed knowledge structure. This idea has been supported by experiments on categorisation (eg, Rosch, Mervis, Grey, Johnson & Boyes-Braem, 1976). It was shown, that in their everyday reasoning people operate better with objects of middle-level abstraction, than with objects of high- or low- level abstraction. For example, people tend to use the category "chair" instead of the more abstract "furniture" or less abstract "kitchen chair." In terms of taxonomic hierarchy, a category that is more abstract is more inclusive, and is thus mainly a superordinate for several subordinate categories. A less abstract category is less inclusive, and is mainly subordinate to several superordinate categories. A category with mid-level abstraction ("basic category" in Rosch's term) is equally associated with both superordinate and subordinate categories. In other words, the well-developed middle-level concepts facilitate fast recognition of a general problem type beyond concrete situations, and help to find the right solutions.

Concept hierarchy is also thought to explain some features of cognitive processing that differentiate experts from novices. In particular, experts categorise problems by their substantive features while novices rely on superficial features (Chi, Feltovich & Glaser, 1981); experts retrieve the required information from memory more easily than novices (Ericsson, Patel, & Kintsch, 2000); and evidently experts find the right solution faster and more often than novices (Larkin, McDermott, Simon, & Simon, 1980).

Based on all these theoretical expectations and empirical findings, we can expect that owing to a more hierarchical structure, highly developed knowledge is better distributed and more easily retrieved. In the next sections, we explore the implications of graph and network theories on CM measurement to approach these characteristics of CMs.

The traditional concept map measures

Generally, a CM is a graphical representation of concepts, represented as nodes and links between them, represented as edges. Traditionally, the number of concepts (nodes) and links between them (edges), together with the concept-links ratio, are used to characterise the quality of knowledge structure in terms of the connectivity. These measures are only objective, which do not require human interpretation. All other scoring systems somehow involve a rater's opinion. For example, in the study of Ruiz-Primo *et al.* (2001), propositions were given points from 1 to 4 based on the accuracy level, while the accuracy was evaluated by raters (teachers). In more traditional scoring

(Novak & Gowin, 1999), propositions are graded based on the level of their qualities, which is also a rater-driven interpretation. Even if the raters are provided with criteria and rubrics for evaluating maps, albeit increasing the reliability, room for subjectivity remains (Watson, Pelkey, Noyes, & Rodgers, 2016). Another problem with traditional measures is their high subject- and theme-specificity, that does not allow researchers to generalise findings across different subject domains.

Graph and network theory for concept map evaluation

During the past decade, researchers have turned to computational theories to evaluate CMs, such as graph theory (Ifenthaler, Masduki, & Seel, 2011; Jamieson, 2012; Stockwell, Smith, & Wiles, 2009; Tyumeneva, Kapuza, & Vergeles, 2017; Zouaq, Gasevic, & Hatala, 2011) and network theory (Frerichs *et al.*, 2018; Koponen & Nousiainen, 2014; Siew, 2018; Stockwell *et al.*, 2009; Zouaq *et al.*, 2011). These new approaches have used information indices to evaluate the complexity of CMs (Bonchev & Buck, 2005), and several measures such as the hyperlink-induced topic search (HITS) centrality or PageRank are used to evaluate the connectivity of CMs (Estrada, 2011; Newman, 2018). Being unified over subject domains and independent of experts, these network measures substantially improved the traditional methods of CM evaluation. In what follows, we introduce three compound measures to operationalise the properties of the CMs. The newly proposed compound measures are based on the well-known HITS and PageRank measures, initially devised to explore knowledge retrieval and storage in large databases. These measures alone, however, were not discriminative enough as such, and lacked the resolving power needed for classification of the CMs discussed here. However, suitable combinations of them can be used to form new compound measures, which turn out to have discriminative and resolving power, bringing out the differences in the network structure. Moreover, as will be seen, such compound measures also yield a qualitative interpretation, which makes them theoretically motivated choices for the analysis.

The HITS and PageRank centralities are based on the idea of tracking different available paths to nodes, and are thus directly connected to information flow, storage and retrieval in the networks. These measures are also closely related to the ways through which the knowledge is retrieved, eg, from the World Wide Web (www), and how search engines operate. In brief, PageRank measures how within a directed network, one node can be reached from all other nodes, ie, how other nodes point to it and how important the node is within the networks on the basis of these connections (Brin & Page, 1998; Li & He, 2018). The HITS centrality is closely related to the PageRank, but it considers the balance between the incoming and outgoing links in more detail (Kleinberg, 1999). The HITS centrality is composed of two values called hubs and authorities. The hub value denotes how many nodes a given node refers to or leads to, and as such, the hub value is related to the capability of the node in acting as a storage or compiler for the knowledge. The authority value is related to the number of connections leading to a given node, and reflects the node importance. The HITS thus resolves the role of the node as stored, and distributes the content of the node onto other nodes of the network. A more detailed description of how HITS and PageRank centralities are defined is provided in Appendix A.

As is evident, the HITS and PageRank centralities come close to notions of knowledge storage, distribution and retrieval, but these measures as such turn out to have limited use in small networks such as studied here. The PageRank and HITS centralities measure properties related to the global connectedness of the nodes. In this study, we use both of them because they approach the connectedness of nodes differently. The PageRank is used to explore the retrieval, because retrieval, as understood here, is related to the reachability of the given node. The HITS centrality is used to explore the distributivity, because distributivity is related to the balance between the incoming and outgoing information.

Based on these network measures, we constructed three compound measures that have better resolving power, and that are more closely related to qualitative ideas about how knowledge storage, distribution and retrieval may emerge in a network.

Three new compound measures based on the network approach

Knowledge structure, represented as a CM, has formal characteristics that are common for a certain level of competence. In particular, we expect that characteristics reflecting the complexity, number of hierarchical levels and concept cohesiveness at each level would show differences between experts and novices (students).

We framed our measurement model by network analysis. Since CMs are directed, we used HITS centrality (hubs and authorities) to evaluate the role of concepts in information transmission within the network, and PageRank as a measure of the structural connectivity, and network diameter as an indicator of map size. The network diameter is the maximum length among all the shortest paths between each pair of nodes, and thus measures the extension of the network. Based on these basic measures, we defined three compound[1] knowledge structure measures to characterise how knowledge is stored and distributed in the networks, and how it can be retrieved. These three measures are knowledge storage capacity, knowledge distributivity and knowledge retrieval.

Knowledge storage capacity *S*, as defined in Equation (1), is a normalised measure for the ratio of hub (H) to authority values (A) of the node (see Appendix A), as they are defined by the HITS centrality (Estrada, 2011; Newman, 2018). Nodes with high authority store knowledge and act as repositories. A high hub value means that the node points to high authority nodes. Thus, dividing the hub value by the authority value can be considered a measure of how knowledge is stored in the nodes to obtain the (relative) knowledge storage capacity:

$$S = \frac{H}{A}, \tag{1}$$

where H denotes the mean of the hub values, and A denotes the mean of the authority values, as defined by the HITS centrality (Estrada, 2011; Newman, 2018). The value of S can vary from 0 to 1, where $S = 1$ means balanced storage in the sense that nodes in the network point most optimally to knowledge repositories, while $S = 0$ means that irrespective of the amount of knowledge in the repositories, no other nodes are connected to it, and knowledge in that node is thus not accessible.

The knowledge distributivity *D* describes in a different way how knowledge is stored in the network. It is a measure that describes the global or local hub distribution, and the difference between the hub and authority values of the node. It is thus a compound measure for the distributivity and diversity of the node's role in the network. The knowledge distributivity D is defined as a logarithmic measure, given by the following:

$$D = \log(T * H(1 - A)), \tag{2}$$

where T denotes the diameter of the networks, and thus a measure for average extension of connections within the network. The value H denotes the mean hub values (from HITS centrality), and A denotes the mean of authority values (from HITS centrality). To interpret the measure, we remind that the nodes with high knowledge storage $S \rightarrow 1$ will have high values of Aà1, indicating thus low distributivity of $D << 0$ (no diversity). Note that the logarithmic form is chosen for practical reasons to provide values that can be easily compared.

The knowledge retrieval R describes how easily knowledge can be retrieved from a network by starting from any of the nodes within it. The retrieval is affected by the knowledge storage capacity S, and by how easily a given node can be reached in the network. To describe this latter property, the PageRank of a given node is a useful value. Consequently, we define the knowledge retrieval R as the geometric mean given by

$$R = \sqrt{S * P}, \tag{3}$$

where S is the knowledge storage capacity and P is the mean of the PageRank of all nodes.

In what follows, we show how the three compound measures, as defined by Equations (1)–(3), provide a good resolving power to find relevant differences between novice and expert CMs. It should be noted that the compound measures S, D and R are based on the HITS- and PageRank centrality, and that diameter T is based on the exact counting of the links, without any human interpretation. The reason for using the compound measures is simply the a posteriori notion that they bring forward differences between the CMs, while basic measures such as HITS and PageRank do not manage to it. Therefore, the choice of using the compound measures S, D and R is practical one, and through interpretation, they can characterise the knowledge storage, distribution and retrieval. Although this interpretation is intelligible, it should not be taken too literally since it is based on theoretical notions about how the network represents these knowledge characteristics, and not on the empirical validation of such representation. Nevertheless, with these restrictions, they can be used as useful conceptualisations in exploring CMs.

Current study

The goal of this study was to explore how the newly developed measures can detect learning-related changes in CMs. Accordingly, we compared CMs of the same students when they were at the very beginning of their semester-long introductory statistics course, and when they completed it. Additionally, as a criterion group, we used professionals in statistics who also produced their CMs (expert CMs). The expert CMs served as a benchmark to evaluate changes in the student CMs. Theoretically, we could hypothesise that a CM changes if it reflects real changes in the knowledge structure, is primarily related to the complexity and connectivity of the CMs, stores concepts, and is available for retrieval. All these changes are supposed to be reflected by relevant CM measures.

We searched for the changes by contrasting: (1) CMs of the same students between the initiation (beginner-level student CMs) and completion of a statistics course (trained student CMs), (2) beginner-level student CMs and expert CMs and (3) trained student CMs and expert CMs.

Our expectations were quite specific. To test them, we used the one-sided t-test[2], choosing hypothesis according to theoretical expectations. Namely, all the newly developed measures as well as the concept degree and mean hubs, are expected to be higher in the trained students, and especially in experts in comparison with beginner-level students. Conversely, the concepts to propositions ratio, mean authorities and the mean PageRank, are expected to be lower in these groups, since it standardised.

We used other indicators for CMs to validate our new measures. We used already known network measures, the HITS centrality and the PageRank, which are expected to be higher in trained students and experts, than in the beginner-level students. We also used the traditional measures, the number of concepts and prepositions, their ratio, and the average concepts degree. Since, the ratio is the number of concepts divided by the number of propositions, we expected that trained students and experts might have a lower ratio than beginner-level students (because the number of propositions, being a denominator, will decrease the ratio).

Data

Ten first-year master's students studying statistics and four professors in statistics were recruited for this study. Each student constructed two maps during the course "Statistical data analysis."[3] No students had any experience with statistical data analysis before this course. The first maps (beginner-level student CMs) were constructed in 6 weeks after this course began. The second maps (trained student CMs) were constructed at the end of the course 6 months later, before the final exam.

Four professors (experts) in statistics also constructed CMs.[4]

The procedure was the same for the students and the professors. It included standardised written instructions to draw a CM on the topic "Statistical Data Analysis" (Appendix B).[5] Respondents became acquainted with the instructions individually, and then started to draw their maps. In most cases, drawing took 20–45 minutes, but the time was not limited. Respondents had an opportunity to consult with the experimenter about the procedure.

Results

The results section is organised as follows. First, we report data from traditional measures used for CM analysis to verify our findings. Second, we report data from more complex network analysis measures. Third, we report how the three newly developed measures (knowledge storage capacity, knowledge distributivity and knowledge retrieval) work for different groups.

1. Traditionally used CM indicators

Figure 1 shows that there was a large variation in the number of propositions and concepts within each group. Experts had more propositions (21.3 in average) in their maps while having the same number of concepts with students. To show this relation, we calculated the concepts to propositions ratio (Figure 2). If the ratio value was larger than 1, the number of concepts was greater than the number of propositions, which shows a low level of the map's connection. The ratio is significantly larger for students than for experts ($t(12) = -6.66$, $p < .01$ for the first measure and $t(12) = -9.70$, $p < .01$ for the second). In other words, the experts' maps were more connective (mean = 0.61) than the students' maps (mean = 1.02 for beginner-level and 1.09 for trained students). There were also significant differences between the students in the two measures ($t(9) = -2.39$, $p < .05$), but trained students had a ratio larger than in the beginning of the course, which is the opposite of what was expected. The same trends were found in the average concepts' degrees ($t(12) = 5.28$, $p < .01$ for the expert and the first measure, $t(12) = 6.45$, $p < 0.01$ for the second), except insignificant difference between students ($t(9) = -0.26$, $p = .39$) (Figure 3).

2. Traditionally used network analysis indicators

The hubs statistic is based on how many outgoing propositions a concept has in terms of the structure, and the possibility to transmit the information (Figure 4). Conversely, the authorities statistic is based on the incoming propositions (Figure 5). These statistics were standardised

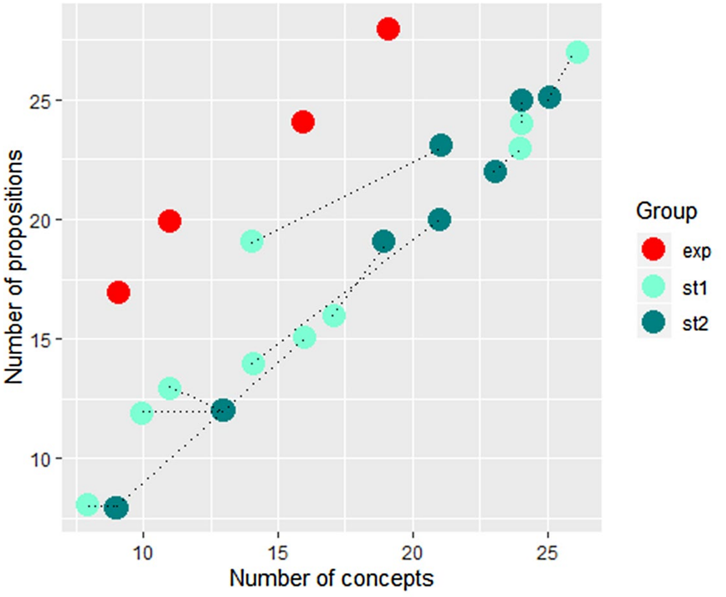


Figure 1: Relation between the number of propositions and the concepts in groups with a dynamic (Hereinafter: "exp"–experts, "st1"–beginner-level students, "st2"–trained students)

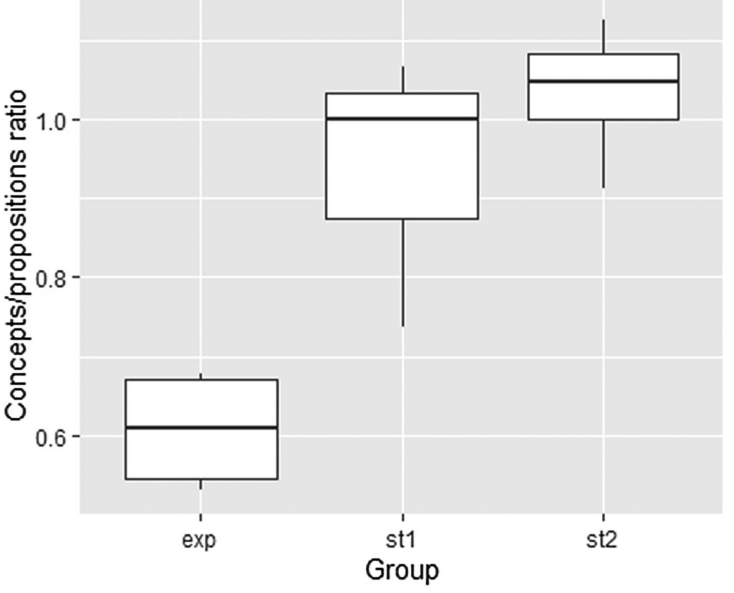


Figure 2: The concepts to propositions ratio in groups (The box plot demonstrates the quartiles of the distribution: the first, the second, and median between them.)

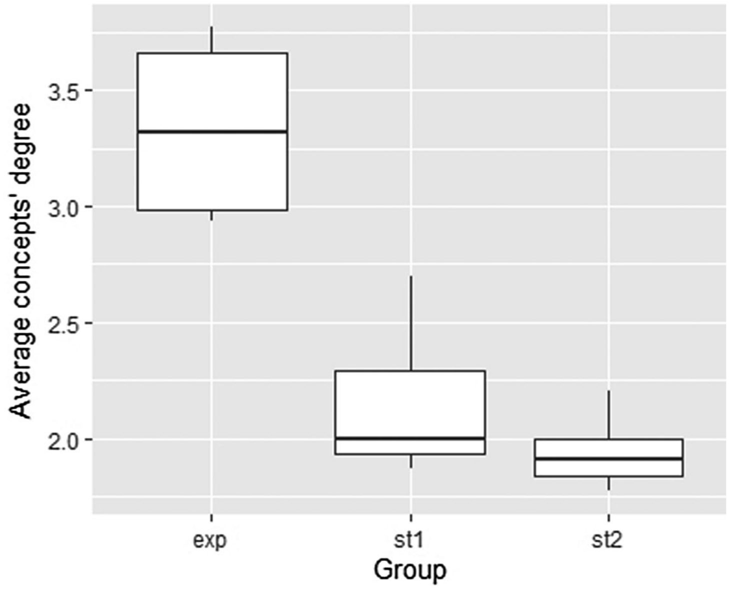


Figure 3: The average concepts' degree in groups

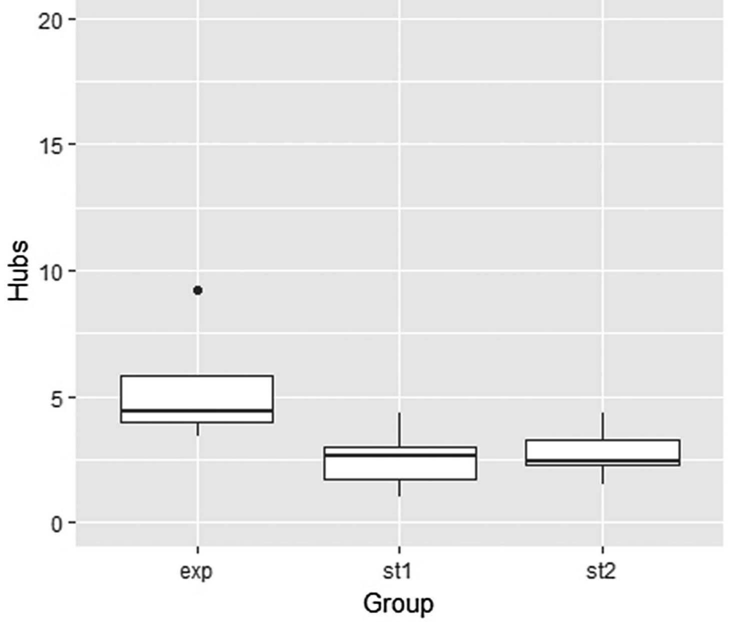


Figure 4: Values of hubs statistic in groups

for comparison. The difference in the hubs statistic was significant for experts and students in both measures ($t(12) = 2.12$, $p = .05$ and $t(12) = 1.94$, $p < .1$), as well as in the authority ($t(12) = -4.01$, $p < .01$ and $t(12) = -4.78$, $p < .01$). At the same time, we did not see significant

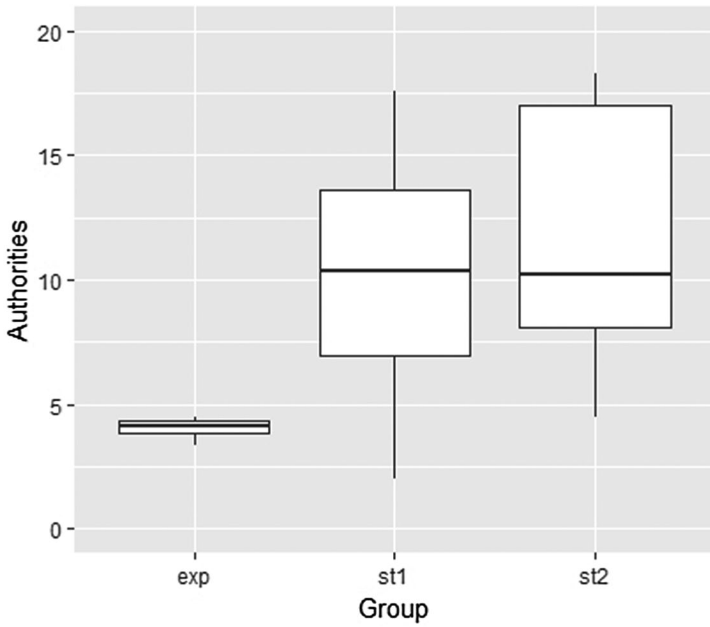


Figure 5: Values of authorities statistic in groups

changes for students between the first and the second measures ($t(9) = 0.65$, $p = .26$ for the hubs and $t(9) = 0.91$, $p = .80$ for the authorities).

Regarding the relationship between these two values, the experts demonstrated almost the same number of hubs and authorities, which means that their networks were balanced. At the same time, students in both measures showed a low value of hubs and a high value of authorities, which means that in their networks, concepts received the information more often than transmitting it (Figure 6). Thus, these indicators support and extend classical CM indicators, such as the concepts to propositions ratio and the average concepts' degree.

The PageRank statistic, in addition to the HITS centrality, shows the possibility of information transmission (Figure 7). Experts had a lower PageRank value, so the importance of each concept in a network is almost the same $t(12) = -1.52$, $p <.1$ between the experts and the first measures of the students, $t(12) = -2.09$, $p < .05$ between the experts and the second measures. Students had a large value for this statistic, which means that there were concepts that stored the information. Although differences between the two measures are visible and predicted, they were insignificant: $t(9) = 1.08$, $p = .84$.

3. Newly developed indicators

As expected, the experts had a knowledge storage capacity that was higher than the students ($t(12) = 4.01$, $p = .01$ and $t(12) = 4.13$, $p = .01$) (Figure 8). Moreover, their results, with a single exception, were close to 1 (mean = 0.85 for experts, 0.26 for beginner-level, and 0.25 for trained students), which implies a rather balanced storage in the sense that nodes in the network point most optimally to the knowledge repositories. In terms of the knowledge distributivity, the experts also obtained higher results than the students during the first ($t(12) = 4.22$, $p < .01$) and second ($t(12) = 3.57$, $p <.01$) measures (mean = 3.04 for experts, 1.84 for beginner-level, and 1.96 for

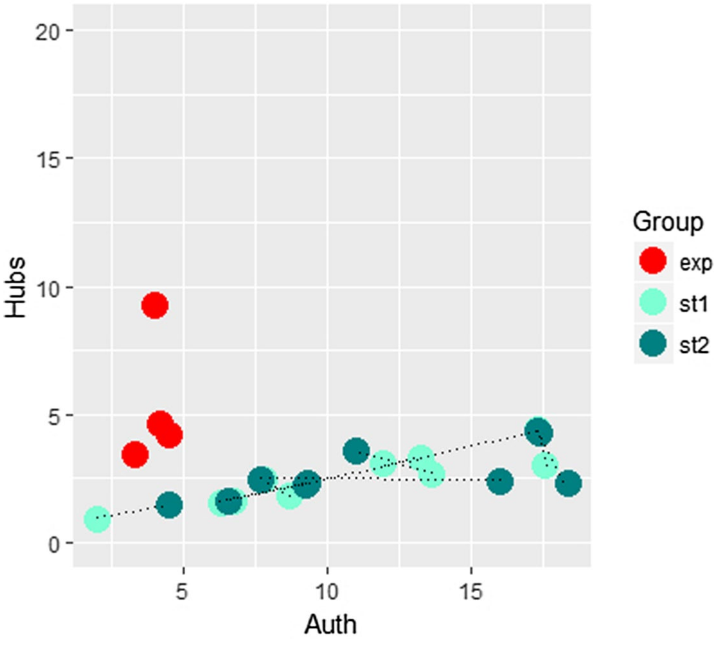


Figure 6: Relation between the hubs and authorities statistics in groups with a dynamic

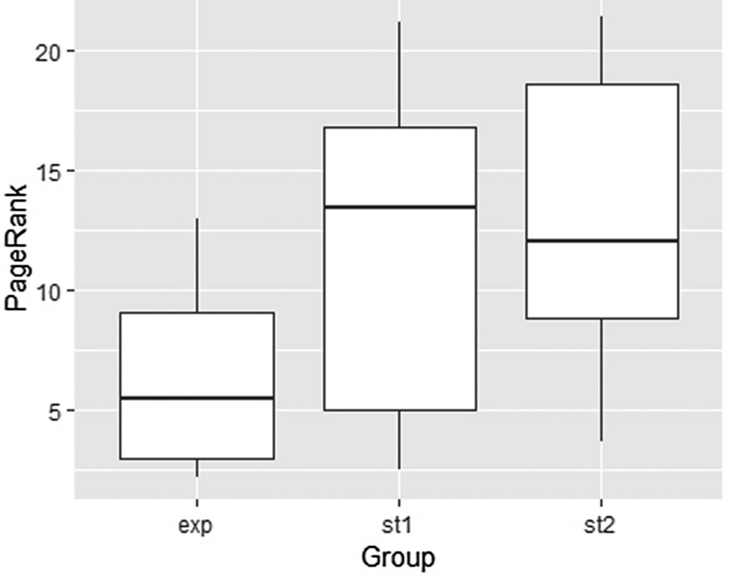


Figure 7: Values of the PageRank statistic in groups

trained students) (Figure 9). The difference between the groups in knowledge retrieval was less than in other indicators (mean = 2.10 for experts, 1.62 for beginner-level, and 1.70 for trained students) (Figure 10). However, experts still had a higher level of knowledge retrieval, according

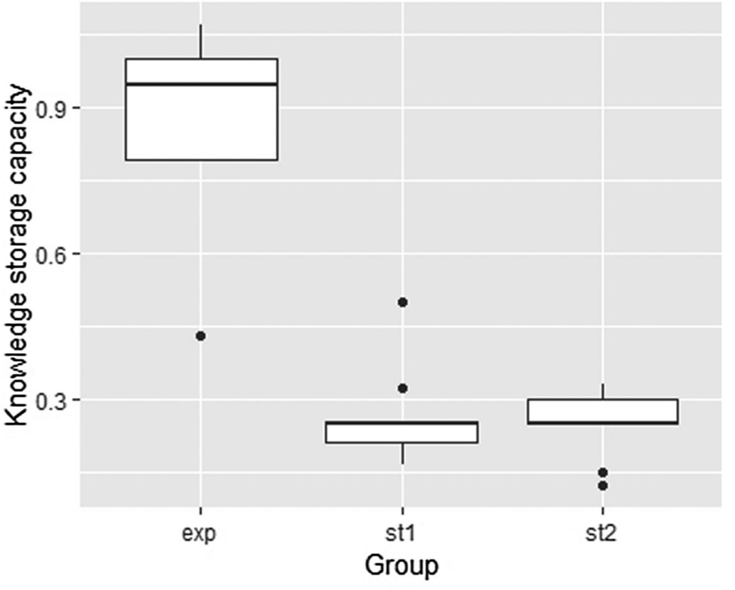


Figure 8: Values of knowledge storage capacity statistic in groups

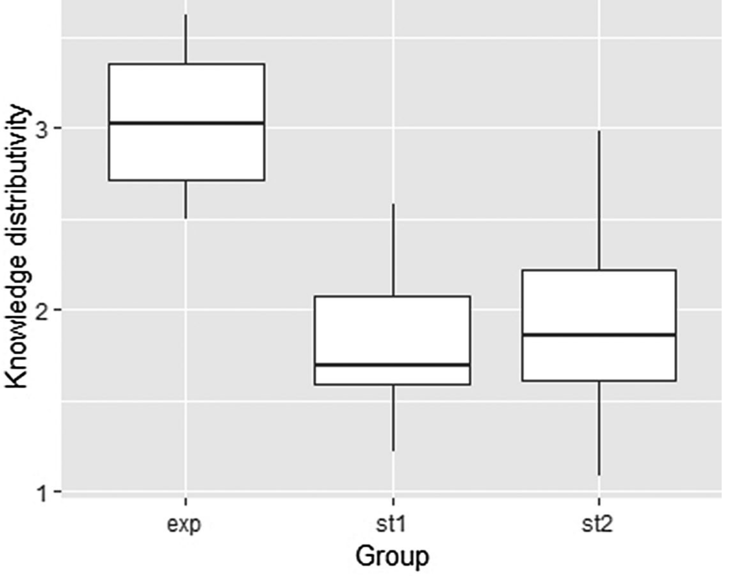


Figure 9: Values of knowledge distributivity statistic in groups

to our expectations, although insignificant ($t(12) = 1.29$, $p = .11$ and $t(12) = 1.18$, $p = .15$). It is important to note a slight but clear increase of two indicators for students between the two measures ($t(9) = -0.38$, $p = .64$ for the knowledge storage capacity, $t(9) = 0.71$, $p = .24$ for the knowledge distributivity, $t(9) = 0.46$, $p = .32$ for the knowledge retrieval). However, the effect size was too small to be detected via parametric and nonparametric tests.

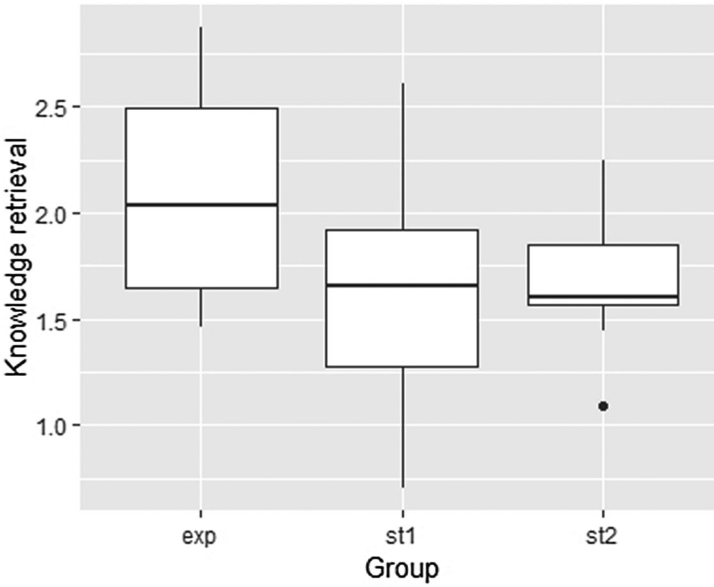


Figure 10: Values of knowledge retrieval statistic in groups

Discussion and conclusions

We planned to show that the knowledge structure represented in a CM has formal characteristics available for objective evaluation, which is common for a certain level of competence. In particular, we expected that characteristics reflecting the complexity, levels of concept abstraction and concept connectivity would differ between experts and novices.

We tested three groups of indicators. First, there were traditional CM measures. These are related to the complexity and connectivity of a CM, measured via a number of concepts such as the concept-to-proposition ratio and the average concepts' degree. In line with many other studies with CMs, these measures were higher in the experts than in the novices (both groups of students), showing that our groups of experts and novices were truly different groups, and could credibly be used to test other measures from network theory that we intended to test. However, we found a negative shift in the concepts propositions ratio for students during the course. Given that this ratio is a measure of the connectivity, we should conclude that students' maps became less coherent during learning. It can be attributed to the rising number of learned concepts, which were still not well interconnected. This result clearly demonstrates the need for alternative approaches to connectivity measure.

The second group of measures was from the network analysis. They are related to the outgoing and incoming propositions and their ratios. The experts' CMs demonstrated that their knowledge structures were balanced in terms of this ratio, while the novices' concepts more often received the information than transmitted it.

However, the most valuable findings are regarding the third group of indicators. The two new compound measures, knowledge storage capacity and knowledge distributivity, revealed significant differences between the experts and novices, with higher values for the experts. It proves that expert knowledge is better distributed, and has a more developed and balanced structure.[6] The difference we found between the novice and expert concept networks is in strict accordance with the theories of cognitive development, as well as with previous experimental findings about features of knowledge structure in experts and novices. An increasing value of the knowledge storage capacity means that experts have more balanced knowledge structures—each node has similar values of authorities and hubs. In other words, the novices mostly have one or two concepts of very high importance and many concepts of very low importance; while experts mostly have concepts with similar importance. Increasing value of the knowledge distributivity means that experts have better distributed concepts within the structure than novices, controlling the size of the structure. In other words, novices mostly operate with concepts that are more local and more independent from each other, while experts operate with concepts which are distributed over the network more globally, without segregated groups of concepts. The third new measure, knowledge retrieval, still showed a higher level of knowledge retrieval in experts than in novices, although the differences were insignificant. This insignificance may be explained by how the knowledge retrieval measure is calculated. The square root of the product of the PageRank and HITS centrality reduces the differences obtained between novices and experts in these latter measures; that is, getting a significant difference in knowledge retrieval would require much stronger novice-expert differences in their knowledge capacity and distibutivity than we obtained.

What is possibly even more important is that these compound measures were sensitive enough to show some progress for the students, albeit statistically non-significant, while common indicators from network theory demonstrated even controversial shifts. The fact that the changes in concept availability were not significant can be explained by the interval between the two occasions of assessment (6 month gap), which might have been too short to allow structural changes to become established. Generally, CM analysis seems to benefit from the network approach in that it provides more sensitive measures, while being in line with traditional analysis in revealing the differences in general.

As mentioned before, the advantages of the holistic scoring system are its capability to evaluate the quality of the structure as a whole, while the traditional approach allows us to make more objective evaluations of individual map elements. The approach developed in this study enables us to bridge the gap between these two, providing an objective evaluation of a CM structure while being sensitive to learning-related changes in the knowledge structure.

It is worth noting that we have developed our measures as a technical tool to scan CMs and extract further information from the combination of single structural values (HITS and PageRank). However, the notion for these compound measures as well as their interpretation were guided by theoretical ideas about changes in knowledge structure during learning and cognitive development. Specifically, it was theoretically reasonable to expect that the difference of network organisation of the knowledge between experts and novices can be described in terms of its retrievability and distributivity across the network. These theoretical grounds made the empirically discovered differences justifiable, and showed that our measures scan networks appropriately. Nevertheless, further research is definitely necessary. The validity and generalisability of our measures are still questionable due to the small sample size, which was also restricted to a single knowledge domain. Strictly speaking, we cannot make a conclusion about the measures' sensitivity to competency growth, and about their cross-domain validity without additional studies on larger samples. However, even in such a small sample, we can still see significant differences between experts and novices, which means that the proposed measures are powerful enough. It would be especially important to set thresholds for the newly developed measures to maximise their usefulness as objective assessment tools. In this sense, the current findings, albeit encouraging, should be considered as rather preliminary.

Funding information

A. Kapuza has worked under the article within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) and supported within the framework of a subsidy by the Russian Academic Excellence Project "5-100".

Statements on open data, ethics and conflict of interest

Please contact the corresponding author to request access to research data.

The study was undertaken in line with Institutional ethics procedures and guidelines. All participants were guaranteed anonymity. According to national regulations, no decision by an ethical committee was needed.

The authors have no conflicts of interest to disclose.

Notes

[1]The compound measures are averages over the more fundamental measures, and thus weight differently eg, aspect of storage and distribution. Their justification derives from their practical success in differentiating between nodes in the network. Such definition of compound measures is common, eg, informetrics and scientometrics (Chen *et al.*, 2009).

[2]We used the Welch approximation for unequal variances. To verify our results, we used the Mood's median test. The Mood's test showed the same results as the *t*-test, so we decided not to include these results in the paper. All computations were conducted in R software.

[3]The course was based on traditional methods of learning: students attended lectures and then discussed topics at seminars; and in addition, students individually worked on data analysis at home.

[4]Examples of beginner-level student and expert maps are in Appendix C.

[5]There were three CMs with a list of concepts provided for students and seven CMs without. This was a part of another project, and the differences between instructions did not affect the current results.

[6]We added examples of beginner-level student maps and expert maps to illustrate differences in their knowledge structure in a more concrete and visible way than via the network measures (Appendix C).

[7]For more details see https://reference.wolfram.com/language/ref/PageRankCentrality.html

[8]For more details, see https://reference.wolfram.com/language/ref/HITSCentrality.html

References

Bonchev, D., & Buck, G. A. (2005). Quantitative measures of network complexity. In D. Bonchev & D. H. Rouvray (Eds.), *Complexity in chemistry, biology, and ecology* (pp. 191–235). Boston, MA: Springer. https://doi.org/10.1007/0-387-25871-X_5

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, *3*(3), 191–209.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Ericsson, K. A., Patel, V., & Kintsch, W. (2000). How experts' adaptations to representative task demands account for the expertise effect in memory recall: Comment on Vicente and Wang (1998). *Psychological Review*, *107*(3), 578–592. https://doi.org/10.1037/0033-295X.107.3.578

Estrada, E. (2011). *The structure of complex networks: Theory and applications*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199591756.001.0001/acprof-9780199591756

Frerichs, L., Young, T. L., Dave, G., Stith, D., Corbie-Smith, G., & Hassmiller Lich, K. (2018). Mind maps and network analysis to evaluate conceptualization of complex issues: A case example evaluating systems science workshops for childhood obesity prevention. *Evaluation and Program Planning*, *68*, 135–147. https://doi.org/10.1016/j.evalprogplan.2018.03.003

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. *Instructional Science*, *39*(1), 41–61.

Jamieson, P. (2012). Using modern graph analysis techniques on mind maps to help quantify learning. *Frontiers in Education Conference Proceedings*, *2012*, 1–6. https://doi.org/10.1109/FIE.2012.6462222

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632. https://doi.org/10.1145/324133.324140

Koponen, I. T., & Nousiainen, M. (2014). Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge. *Journal of Complex Networks*, *2*(2), 187–202. https://doi.org/10.1093/comnet/cnu003

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*(4450), 1335–1342.

Li, K., & He, Y. (2018, April). Evaluating nodes importance in complex network based on PageRank algorithm. In *AIP Conference Proceedings* (Vol. 1955, No. 1, p. 040122). AIP Publishing LLC. https://doi.org/10.1063/1.5033786

Newman, M. (2018). *Networks* (New Edition, 2nd ed.). Oxford, UK: Oxford University Press.

Novak, J. D., & Gowin, B. (1999). *Learning how to learn* (reprinted). Cambridge, UK: Cambridge University Press.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: W.W. Norton & Co. https://doi.org/10.1037/11494-000

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, *38*(2), 260–278.

Siew, C. S. Q. (2018). Using network science to analyze concept maps of psychology undergraduates. *Applied Cognitive Psychology*, *33*(4), 662–668. https://doi.org/10.1002/acp.3484

Stockwell, P., Smith, A. E., & Wiles, J. (2009). Displaying a framework in a concept map using network graph techniques. In *2009 13th International Conference Information Visualisation, Barcelona, 2009* (pp. 661–666). https://doi.org/10.1109/IV.2009.36

Tyumeneva, Y., Kapuza, A., & Vergeles, K. (2017). Distinctive ability of concept maps for assessing levels of competence. Pilot study. *Voprosy Obrazovaniya/Educational Studies Moscow*, *4*(2017), 150–170. https://doi.org/10.17323/1814-9545-2017-4-150-170

Vygotsky, L. (1982). *Myshlenie i rech* (Vol. 2). Moscow: Pedagogika.

Watson, M. K., Pelkey, J., Noyes, C. R., & Rodgers, M. O. (2016). Assessing conceptual knowledge using three concept map scoring methods. *Journal of Engineering Education*, *105*(1), 118–146. https://doi.org/10.1002/jee.20111

Zouaq, A., Gasevic, D., & Hatala, M. (2011). Ontologizing concept maps using graph theory. In *Proceedings of the 2011 ACM Symposium on Applied Computing, Taichung* (pp. 1687–1692). https://doi.org/10.1145/1982185.1982537

Appendix A

PageRank and hits centralities

The PageRank centrality measures the importance of a node on the basis of how different nodes with different importance refer (are connected) to it; the more the number of important nodes referring to a node, the higher is the PageRank centrality of that node. The PageRank centrality was originally introduced by Brin and Page (1998) for identification of important webpages in World Wide Web (WWW). The PageRank centrality is defined iteratively (recursively) based on how all nodes within the network are connected. Consequently, a node, which is connected to many nodes with high PageRank centralities receives a high PageRank centrality. To simplify, the PageRank computes the importance of a particular node proportionally to the importance of the other nodes pointing to this particular node. This algorithm is successfully used for web page ranking in Google search, putting first more important (received more links from other websites) websites.

For example, one node may be directed to by numerous nodes with low importance; another node may be directed to by only a couple of nodes with high importance. In the final importance of the first and the second nodes, their PageRanks will take into account both the number and the importance of the references.

Technically, the PageRank centrality is based on incoming links, and expresses the probability that a given page is reached from through random walk. The probabilities are assigned values from 0 to 1, a certain value meaning the probability of the given node being reached by random exploration of the network (0 means a node is never reached, and 1 means a node could be reached from any point). Roughly, large value of the PageRank means that a node is easily found by traversing the network (even by a random search) and easily reached, while a low value indicates that it is difficult to find and reach. Such probabilities are found through iterative exploration of the networks, where the PageRank values are iteratively updated (Brin & Page, 1998; Newman, 2018). First, each node is assigned the same value, and then (by iterations) the value of each node is distributed to all the neighbour nodes, and the new value of each node is the sum of the obtained values (Li & He, 2018). However, there is no simple closed form expression or matrix-resolvent on how to calculate the PageRank exactly. In the present study, the PageRank is obtained by using MATHEMATICA library and its function PageRank Centrality for calculation of the PageRank[7].

The HITS centrality, introduced by Kleinberg (1999), is in many respects similar to the PageRank, but it differentiates between incoming and outgoing links. The basic idea behind the HITS centrality is the notion that certain nodes act as directories or repositories of information or knowledge, and point to many other nodes where that information or knowledge is channelled. Such nodes are called hubs. In addition, certain nodes may receive many links from the hubs; such nodes are called authoritative. Consequently, the HITS centrality consists of two values, hubs and authority. Therefore, a node has a high hub value if it has many outgoing links to nodes with high authority values; a node has a high authority value if it has many incoming links from nodes with high hub value.

Technically, the hub and authority values of HITS centrality are calculated iteratively, through mutual (reciprocal) recursion. The authority value of a node is obtained through the hub values pointing to that node, whereas the hub value is obtained by summing the authority values of the nodes the given node is pointing to (Kleinberg, 1999; Newman, 2018). A hub value depends not only on the number of nodes that it points to, but also on the authority of those nodes; and the same is true for the authority value. The algorithm firstly calculates each node's authority value as the sum of the hub values (which are initially 1) of each node that points to it. Second, each node's hub value is recalculated as the sum of the authority value of each node to which it points. Then, each value is normalised, and the algorithm repeats. In this study, the HITS centrality has been calculated using the MATHEMATICA library and its function HITS Centrality[8].

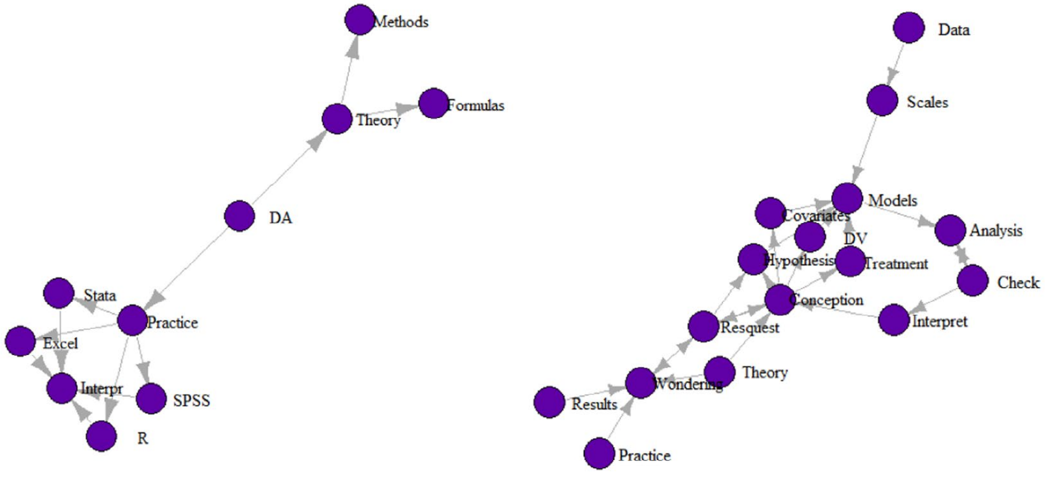
Appendix B

Instruction

We thank you for participating in the study. Please read carefully the instructions for constructing concept maps. Using an A4 paper and your pencil, draw a concept map of the *Statistical Data Analysis* area. You can use any terms and concepts and combine them as you see fit. Do not try to cover all topics related to statistics, but reflect the key, in your opinion, elements necessary for understanding what *statistical data analysis* is, what it consists of and how it is implemented.

Appendix C

An example of CMs values



	Beginner level student	*Expert*
Number of concepts	10	16
Number of propositions	12	24
Ratio	0.83	0.67
Average degree	2.40	3
Hubs	1.67	4.64
Authorities	6.67	4.24
PageRank	2.46	3.21
S	0.25	0.91
KD	1.50	3.62
KR	0.78	1.71