Finding the Key Concepts of Students' Knowledge
A Network Analysis of Coherence and Contingency of Knowledge Structures

Ismo T. Koponen and Maija Nousiainen
Department of Physics, University of Helsinki, Helsinki, Finland

Keywords: Conceptual Structures, Coherence, Networks.

Abstract: The desired outcome of learning science is students' expert-like subject knowledge, which is expected to be at the same time well-organized, coherent and contingent. However, it has proved difficult to find ways to represent these features and to identify the key conceptual elements or concepts that are responsible for them. In this study concept networks constructed by physics students' representing their views of the relatedness of physics concepts are analyzed in order to clarify how coherence and contingency can be captured and measured. The data consist of concept networks (N=12) constructed by physics students, representing relationships between physics concepts of electricity and magnetism. The networks are first analyzed qualitatively for their epistemic acceptability. The structure of the concept networks is then analyzed quantitatively using a network graph theoretical approach. The analysis picks out a handful of key concepts which all play a central role in all of the concept networks examined. From the physics point of view these key concepts are relevant ones (most of them having to do with fields), which indicates the relevance and power of the method in describing knowledge structures.

1 INTRODUCTION

Expert-like knowledge can be characterized as well organized and utilizable; such knowledge is coherent and contingent at the same time (Derbentseva et al., 2007; Koponen and Nousiainen, 2013). Coherence is obviously connected to structural organization of knowledge (BonJour, 1985; Thagard, 2000), but exactly what such coherence might mean and how to recognize it, is rarely discussed. Contingency of knowledge refers to different and alternative ways of introducing concepts by using the support of already known concepts (Scheibe, 1989). In order to study the coherence and contingency of students' knowledge some suitable representational vehicles are needed to illustrate the relationships obtaining between concepts. One tool for representing pre-service teachers' conceptual understanding is provided by concept networks (Koponen and Nousiainen, 2013; Nousiainen, 2013; Börner et al., 2009).

The question of the structure of knowledge as it is represented in a concept network is related to principles of map design. A concept map which is meant to be an expression or representation of epistemically justified knowledge, as well as communicable to others through argumentation, needs specific rules and certain norms (Nousiainen, 2013 and references therein). Such concept networks provide a lot of information on how students conceive the structure and content of subject matter knowledge, and are in fact related to question how learner construct ontologies in learning. By visual inspection only, however, it has proved very difficult to quantify the essential differences (Koponen and Nousiainen, 2013).

We present here an analysis of concept networks, which takes into account the epistemic justification of knowledge as it is represented in the networks, and which also pays attention to the structure of the networks. As a consequence, coherence and contingency as epistemic and structural notions become defined and yield to operationalization. The method that combines qualitative analysis with detailed quantitative methodology is advantageous in the study of large, connected sets of conceptual elements.

We use here the context of electricity and magnetism as a specific example to show how coherence and contingency emerge as a special type of connectedness or relatedness of concepts, and how the key concepts provide this coherence and contingency. The research questions posed and

Koponen, I. and Nousiainen, M..
Finding the Key Concepts of Students' Knowledge - A Network Analysis of Coherence and Contingency of Knowledge Structures.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 2: KEOD*, pages 239-244
ISBN: 978-989-758-158-8

answered in the present study the following:

- How can the key concepts providing coherence and contingency be identified in students' representations of their knowledge?
- What are the key concepts providing coherence and contingency in the case of electricity and magnetism?

These questions are answered using a sample of concept networks and written supplementary reports produced by physics students', third year university level. The analysis is based on a combination of qualitative and quantitative methods. Qualitative analysis is used in assessing the epistemic acceptability of knowledge expressed in the networks and written reports. Quantitative analysis based on network theory is then used to identify the key concepts providing coherence and contingency.

2 COHERENCE, CONTINGENCY AND KEY CONCEPTS

The question of the organization of knowledge is closely related to the ways conceptual knowledge is acquired and justified using existing conceptual knowledge and existing concepts. The ways that concepts are used in that process tie them together, provide meaning and eventually lead to an interwoven web of concepts wherein they are related. Coherence of such a web of knowledge arises from the mutual support of relations and from the epistemic justification of such relations. In addition, for coherence to be useful and interesting, the knowledge system must also be contingent (Scheibe, 1989; BonJour 1985).

The coherence we are interested in here is the coherent relatedness of concepts and other possible elements of conceptual knowledge, for example models (cf. BonJour, 1985; Thagard, 2000). These coherent relations are based on specific types of situations: using concepts either in the context of describing or explaining the outcomes of experiments, or using concepts as parts of models which describe or generalize experimental results (see Nousiainen 2013 and references therein). Coherence with regard to experiments and experimental observations ensures that a conceptual system can be used in giving explanations and making predictions of observed features of real systems. The use of concepts is systematic and symmetric in the sense that concepts retain their mutual dependencies and relations in different situations (BonJour, 1985; Thagard, 2000).

In educational settings, coherence is established through instruction and argumentation, rather than through genuine discovery. In physics education, instructional settings providing means of introduction of new concepts are most often different kinds of laboratory experiments or modelling activities. The experiments discussed here cover laboratory experiments and the explanations which are given to data produced in such experiments. The models and modelling of relevance here is the most common way to use models in physics teaching, namely providing explanations and predictions (see Nousiainen, 2013, and references therein).

Coherence which is produced through the above-mentioned use of experiments and models connects concepts to each other symmetrically so that if concept A is connected to B, and B to C, a connection between A and C also becomes established. In practice, these types of connections give rise to cyclical basic patterns, of which a 3-cycle of three concepts is the most common (Koponen and Nousiainen, 2013; Nousiainen, 2013). Therefore, in what follows, coherence will be operationalized through special counting of such cyclical, mutually supporting connections.

The contingency of knowledge refers to the different possible conceptual paths with which concepts are related to each other successively, thus providing different and alternative ways of introducing concepts using the support of already known concepts. Coherence and contingency are both important aspects of scientific knowledge and are expected to increase when the body of knowledge expands (Scheibe, 1989; Chen et al., 2009). In teaching and learning, contingency answers to the questions of how, and in how many ways, new concepts are introduced and justified on the basis of concepts which have already been learned. This kind of knowledge is an important part of the learner's, conceptual knowledge (Koponen and Nousiainen, 2013; Nousiainen, 2013). Contingency, as a qualitative notion, is therefore related to the multiplicity of ways a given concept participates in connecting other concepts (BonJour, 1985; Scheibe, 1985). Different concepts in the web have thus different epistemic and structural roles in providing coherence and contingency.

The key concepts of the network are those ones which provide the coherence and contingency of the whole conceptual system. Coherence and contingency, however, are notions that refer mainly to structure. Reference to the epistemic content of knowledge is also needed in order to represent reliable knowledge. The key concepts should also have strong epistemic status in the network, recognized from their role in providing epistemically well-justified connections.

3 EMPIRICAL SAMPLE

The context of the research reported here is an advanced-level course for pre-service physics teachers (third or fourth year students) in which electricity and magnetism were discussed. Students were asked to concentrate on their discussions and reflections on the central concepts, laws, models and experiments they thought important in forming a well-organized picture of the content and structure of electricity and magnetism. The design of the concept networks discussed here is based on special kinds of nodes representing the knowledge. The nodes in these networks represent: 1) quantities, 2) laws, 3) models, and 4) experiments. The linking words are describe possible procedures and actions how nodes are connected. Students were required to provide epistemic justification for every node-link-node chain they draw and describe it in a supplementary written report. Other details of the design rules to construct the concept networks are reported elsewhere (Nousiainen, 2013). The empirical sample analyzed in this study consists of 12 such representations. Here, only the final versions are considered because the final stage of the students' understanding of the relatedness of concepts is of interest in finding the key concepts.

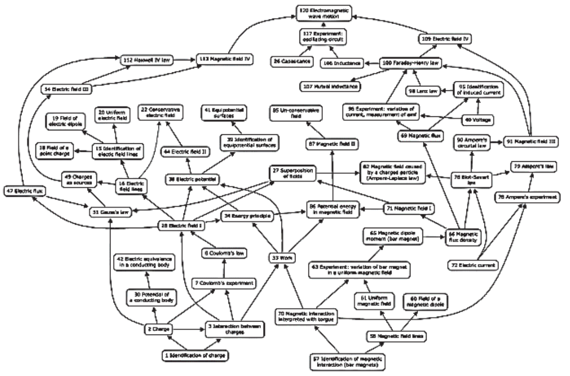

Figure 1: An example of s student's network which contain 55 different concepts. Altogether 121 concepts were found in students' networks. Only schematic, network-like overview is shown, the text in boxes is not meant to be read.

4 ANALYSIS METHOD

The analysis is a combination of qualitative and quantitative methods. The qualitative analysis is carried for epistemic justification of network nodes and links, while the quantitative analysis is used to operationalize coherence and contingency and to identify the key concepts.

The epistemic validity of knowledge concerns the epistemic acceptability of explanations students provide in their written reports. Attention is paid only to following four epistemic dimensions: 1) ontology, 2) facts, 3) methodology and 4) valid justification. These four criteria form a suitable basis for the analysis of epistemic acceptability of knowledge represented in concept networks (Nousiainen, 2013). These criteria form cumulative, hierarchical ladders and values 1-4 are used indicative of the order in which the above epistemic "norms" are fulfilled. In addition to the epistemic analysis of the nodes, also the links were evaluated using similar kind of taxonomy applied to links. The epistemic analysis of written reports and linking words is the only interpretative part of the analysis. It produces the data for the quantitative analysis in the form of epistemic weights of nodes and links in the concept networks.

Coherence and contingency depend on the degree of the epistemic justification and on the overall connections of the system of knowledge, where the key concepts have a special role. In order to find out which concepts are the key concepts, we must operationalize the properties of coherence and contingency. For this, the information contained in the concept network itself must be suitably formalized so that the network yields to quantitative analysis.

Formalization of Networks. After the interpretative analysis, the important information contained in the networks is carried by node and link strengths. All the values were normalized to range from 0 to 1 so that epistemically strong nodes and links have strengths 0.75–1, while weak nodes and links have small values 0.0–0.25. The strength of node i is denoted by s_i, while w_{ij} is the strength of the link from node i to node j. In addition to strength, the node carries a tag τ which specifies the type of the node, either conceptual, experiment or model. The information on epistemic strengths is simplified in further analysis by rescaling link strengths so that the epistemic strength of an initiating node and a link emerging from it are aggregated to form a new weight $w_{ij} \rightarrow s_i\, w_{ij}$, motivated by the notion that directed links w_{ij} pass information from node i to node j. The directional weighting is important, because the ordering of nodes depends on the order of argumentation represented in networks. However, network is symmetrized $w_{ij} = w_{ji}$. For final analysis and direction is taken into account in weights.

Coherence and Contingency Operationalized. The networks are now described fully adjacency matrix W with $[W]_{ij}=w_{ij}$. Coherence is related to closed cycles, and the more there are such cycles a given node (concept) participates in, the larger is the coherence such a node provides for the network. Without further mathematical details we note that coherence can be operationalized as subgraph centrality SC_k (Estrada et al., 2012; Benzi and Klymko, 2013),

$$SC_k = \frac{\left[e^W\right]_{kk}}{\sum_k \left[e^W\right]_{kk}}$$

where k is the given node. Contingency is operationalized similarly by counting open walks between nodes p and q such that a given node k is involved in the walk. The more there are such walks, the more there are contingent paths from p to q supported by k (i.e. more alternatives to connect p and q via k). Contingency is then operationalized as communicability betweenness centrality BC_k (Estrada et al., 2012; Benzi and Klymko, 2013),

$$BC_k = \frac{1}{C}\sum_p \sum_q \frac{\left[e^W\right]_{pq} - \left[e^{W+W'}\right]_{pq}}{\left[e^W\right]_{pq}}$$

where C is a normalization factor $C=(N-1)(N-2)$. The subgraph centrality and betweenness centrality are centrality measures taking into account the whole structure and its connectivity, based on information flow, and thus better suited for purposes of characterizing coherence and contingency than other measures of betweenness and centrality, not directly related to information flow (see Estrada et al., 2012).

The Key Concepts as Importance Ranking. The nodes (concepts) that gain high values of SC_k and BC_k are the key concepts providing the overall coherence and contingency of a concept network. The key can be recognized using a normalized geometric mean of SC_k and BC_k. by introducing importance ranking IR_k (Chen et al., 2009)

$$IR_k = \frac{\left(SC_k BC_k\right)^{1/2}}{\left(\max\{SC_k\}\max\{BC_k\}\right)^{1/2}}$$

Importance ranking ranges from 0 to 1 and does not depend on network size, which makes it possible to make comparisons between very different networks.

5 RESULTS

The sample of N=12 networks had each on average 59 nodes and 97 links. Altogether 121 different concepts were identified, of which 22 were experiments, 37 were models and 72 conceptual nodes. For each node k in each networks we calculated the subgraph centrality SC_k (coherence) and betweenness centrality BC_k (contingency). For comparisons, the strength D_k (number of links) of a node as the sum of weights of links connected to this node is also calculated. The results when all nodes (experiments, models and conceptual) are taken into account are shown in Figure 2.

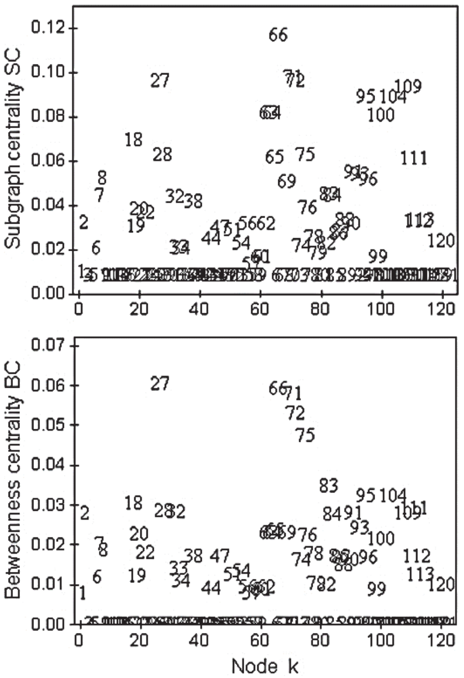

Figure 2: The subgraph centrality SCk (top) the communication betweenness centrality BCk (bottom) for one concept network.

From figure 2 it is seen that a small subset of concepts (labelled 27, 66, 71, 72, 75 and 109) has large values of D_k, SC_k and BC_k; thus they can be identified as concepts that provide coherence and contingency. The measures SC_k and BC_k seem to have power to discern the structurally important concepts but the changes are substantial.

In order to identify the key concepts so that the identification does not depend on the size of the network we calculate the importance rankings IR_k. for each of the student networks separately. Because we are interested to compare individual students' representations, we need to display the data so that each node in each network can be compared. The importance rankings of all 121 nodes, of which about 50-60 appear in a given concept network, can be compared by representing them as a kind of "spectrogram", where most important concepts are shown as black stripes and the least important ones by white stripes. This spectrogram is shown in Figure 3 for all concept networks, with all nodes taken into account, and for nodes with experimental (exp) and model-based (mod) epistemic support separately. The spectrogram makes it possible to compare different networks at one glance. As figure 3 shows (top row), the darkest stripes are concepts 2, 15, 27, 28, 38, 47, 51, 63, 66, 71, 100 and 109, which means that these concepts are important in all concept networks. Almost as important concepts stand out to be 8, 33, 57, 69, 83, 91 and 113. The middle row illustrates that concepts 8, 51 and 57 are mainly supported by experiments, and the lowest row that the concepts 28, 47, 69 and 83 are backed up by models. Note that experiment and model support need to be compared with each other to find out which one dominates. However, most of the important concepts are supported equally by experiments and models and the differences in the importance rankings are not substantial. A summary of the key concepts common for all concept networks is as follows: *charge* (2), *Coulomb's law* (8), experiment: *electric field lines* (15), *superposition of fields* (27), *electric field defined through force* (28), *work* (33), *electric potential* (38), *electric flux* (47), *Gauss law* (51), experiment: *magnetic interaction* (57), experiment: *magnetic flux density* (63), *magnetic flux density* (66), *magnetic flux* (69), *magnetic field defined through torque* (71), *magnetic force of a moving charged particle F=qvB* (83), *magnetic field as independent entity* (91), *Faraday-Henry law* (100), *rotational electric field* (109), *Ampere-Maxwell law* (113).

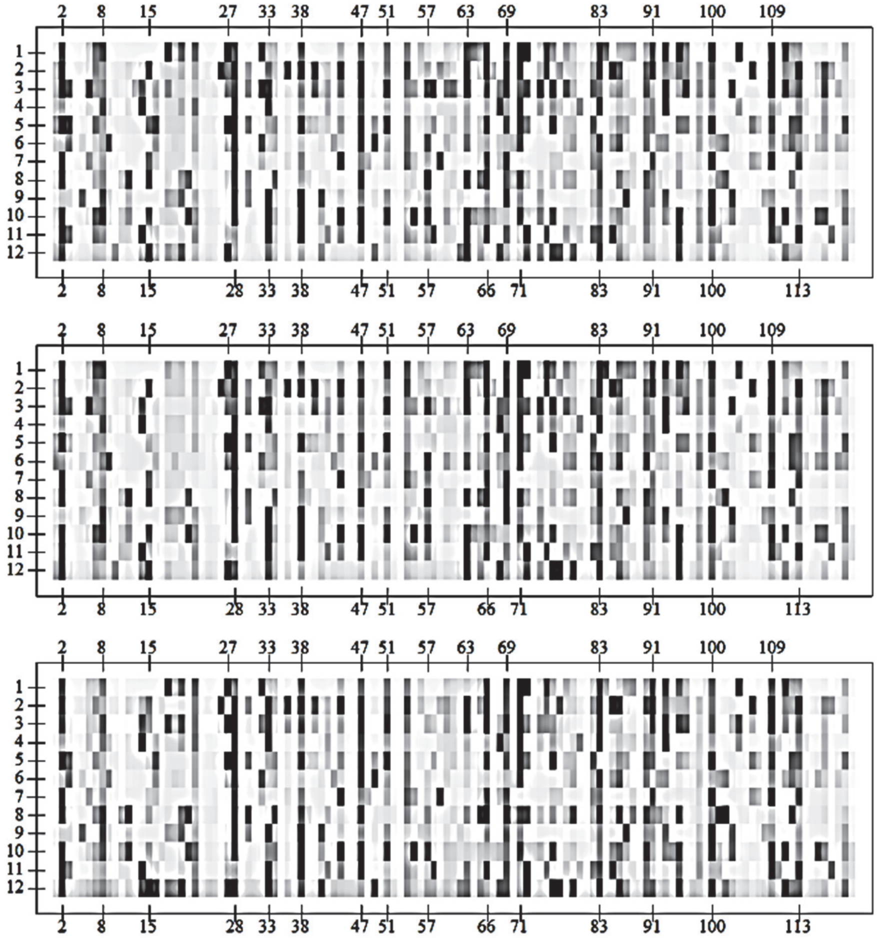

Figure 3: Key concepts as identified on basis of Importance Rankings IR_k of nodes (concepts) for all 12 networks and 121 nodes. Dark stripes denote the highest importance rankings; the lighter the stripe the lower the ranking. Note that certain nodes (denoted in the figure with their numbers) have high rankings in many of the networks. The up-most row is for all concepts, the middle row for experimentally supported and the lowest row for model supported concepts.

The similarity of different networks can now be examined on the basis of the importance rankings. This examination shows that in nearly all concept networks the field concepts (28, 71, 91 and 109) are the most central ones, and for them, experimental support and model support are equally important. If we focus on this core set of key concepts, the networks are similar. In addition to this core set, there is a handful of almost as important concepts (8, 33, 57, 69, 83, 91 and 113) which appear in many of the networks. Although there is much variation between students, there are, however, also many shared key concepts.

6 CONCLUSIONS

Good organization of content knowledge is here approached from the assumption that coherence and contingency are two important qualitative features of well-organized knowledge. These kinds of relations are noted to be central for the functionality of conceptual knowledge (Derbentseva et al. 2007; Koponen and Nousiainen, 2013; Nousiainen, 2013). The method presented here allows us to analyze key concepts which provide the coherence and contingency. Coherence is operationalized through cyclical connections between concepts. Contingency is operationalized as connected, not cyclical, paths between given concepts. The key concepts were found by forming an importance ranking on the basis of these operationalized measures. The importance rankings brought forward a small set of key concepts which have a more important role than other concepts in providing the coherence and contingency for the whole set of concepts. These concepts turn out to be meaningful from the point of view content, too, which is of course a satisfying finding and not trivially expected in this kind of learning context. In all cases epistemic support from experiments and models was found of equal importance, although slightly differently for different key concepts.

Importance rankings also allow us to compare networks: if the same nodes have high importance rankings in two different concept networks, it means that the networks are similar to some extent. The analysis carried out here showed that all the 12 networks inspected here had much similarity in the way they all emphasized the centrality of field concepts.

In summary, our results suggest that concept networks, if properly analyzed, contain valuable information of how students organize conceptual structure in physics. In particular, with network based methods it becomes possible to identify the key concepts that provide coherence and contingency of such concept networks. This kind of knowledge is important to understanding how human learners construct ontologies in learning, how these ontologies may differ, and how learning environments can support the ontology construction by suitable visualizations.

REFERENCES

Benzi, M., Klymko, C., 2013. Total communicability as centrality measure, Journal of Complex Networks, 1, 124-149.

BonJour, L., 1985. *The Structure of Empirical Knowledge*. Cambridge, Massachusetts: Harvard University Press.

Börner, K., Scharnhorst, A. 2009. Visual conceptualizations and models of science. *Journal of Informetrics*, 3, 191-209.

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, 191-209.

Derbentseva, N., Safayeni, F., Cañas, A., 2007. Concept Maps: Experiments on Dynamic Thinking. *Journal of Research in Science Teaching*, 44, 448–465.

Estrada, E., Hatano, N., Benzi, M., 2012. The Physics of Communicability in Complex Networks. *Physics Reports*, 514, 89–119.

Koponen, I. T., Nousiainen, M., 2013. Pre-service Physics Teachers' Understanding of the Relational Structure of Physics Concepts. *International Journal of Science and Mathematics Education*, 11, 325–357.

Nousiainen, M., 2013. Coherence of Pre-service Physics Teachers' Views of the Relatedness of Physics Concepts. *Science & Education*, 22, 505–525.

Scheibe, E., 1989. Coherence and Contingency: Two.

Neglected Aspects of Theory Succession. *Noûs*, 23, 1–16.

Thagard, P., (2000). *Coherence in Thought and Action*. Cambridge, Massachusetts: The MIT Press.