The Semantic Structure of the Slovak Physics Textbook

Petra Králiková[1], Aba Teleki[2]

Abstract

Learning process of the physics without a textbook is unimaginable, so the way of writing the textbook is important. In our work we deal with the curriculum in Slovak physics textbooks from viewpoint of the semantic network structure and then we investigate the modularity of this structure. The self-organized networks (networks created by natural way) show certain universality. These networks have scale invariant distribution (power law distribution) and parameter α has value between 2 and 5. In our textbook analysis we transform the text to a tripartite network and after that we perform network analysis of these networks. We found the same distribution as for networks that are created in a natural way. We have applied cluster analysis on the resulting network and we have acquired individual modules – clusters. These clusters reveal the true connectivity structure of the content and we have examined their properties. The cluster analysis divided the network into 13 clusters, the three largest clusters deal with the three basic state of matter – gaseous, liquid and solid state.

Keywords: *Slovak physics textbooks, semantic structure of network, scale invariant distribution, clusters;*

1. Introduction

Textbooks are very important in learning process, especially physics textbooks. In our work we investigate physics textbooks from the viewpoint of network structures. We have analyzed it by network tools at the level of microstructure. Self-organized networks have in the nature a scale invariant (power law) distribution [1], [2], [3], [4]. The probability *p* of this distribution is

$$p(k) = Ck^{-\alpha}$$

where *k* is the degree of a nod (the number of links it has to other nodes), *C* is determined by the normalization condition $p(1) + p(2) + \ldots = 1$ and the parameter *α* describes the network and it is equal approximately to the average degree of nods. Barabási points that universal properties of self-organized networks are independent on the function, age, or scale of the natural networks and they have roughly the same structures [4]. The network structures of textbooks are still poorly explored. Several contributions are devoted to analyzing the curriculum [5], [6], but not by the way we analyze them.

In our approach, we transform the text of the textbook to a tripartite network, which characterizes the links between three types of nodes (nodes of type sentence, term and formula). The network represents a system which describes interactions between the nodes which are mediated by the links (edges). The network may be described by various network characteristics as number of nodes, number of links, average degree of nodes, shortest path, diameter, average path length, modularity [7]. The universality found in natural networks concerns the value of the parameter *α* of the appropriate scale invariant distribution, and Barabási found that the value of the parameter *α* is between 2 and 5 [8]. It is crucial for networks with scale invariant distribution that many nodes have only one or two links but also there are nodes that have a large number of links.

In our opinion, the microstructure of a textbook (that form a semantic network structure) must be designed to support the inner brain strategy by means of which they learn [9]. We assume that if the text of a textbook is a network with the scale-invariant distribution, it helps for easier and faster learning.

2. Methods

We analyzed the Physics textbook for the 2nd year of high school from 1985, specific topic *Štruktúra a vlastnosti látok* (*The structure and properties of substances*), which represents a 162 page-long educational text [10]. We transform the text into a network consisting of three types of nodes -- a tripartite network (with three types of nodes: sentences, terms and formulae).

[1] Constantine the Philosopher University, Slovak Republic
[2] Constantine the Philosopher University, Slovak Republic

The *sentence type nodes* are sentences, images, tables, notes, tasks, examples (see [11]). The *term type nodes* are physics terms and names of persons. The *formula type nodes* are physics formulae, equations and indications of physical quantities and units.
We consider a word to be a physics term if it is listed in the dictionaries of Physical terms and phrases [12], [13] or in the index of the textbook.
A term or formula, respectively is linked to a sentence when the term or formula is the part of the sentence. A sentence is linked to another sentence if one sentence refers to the other sentence by transitional words or transitional phrases.
We analyzed the mentioned textbook using the tool Gephi.

3. Results

We obtained the tripartite network shown in the Fig. 1, where yellow nodes are *sentences*; red nodes represent terms and blue nodes are formulae.

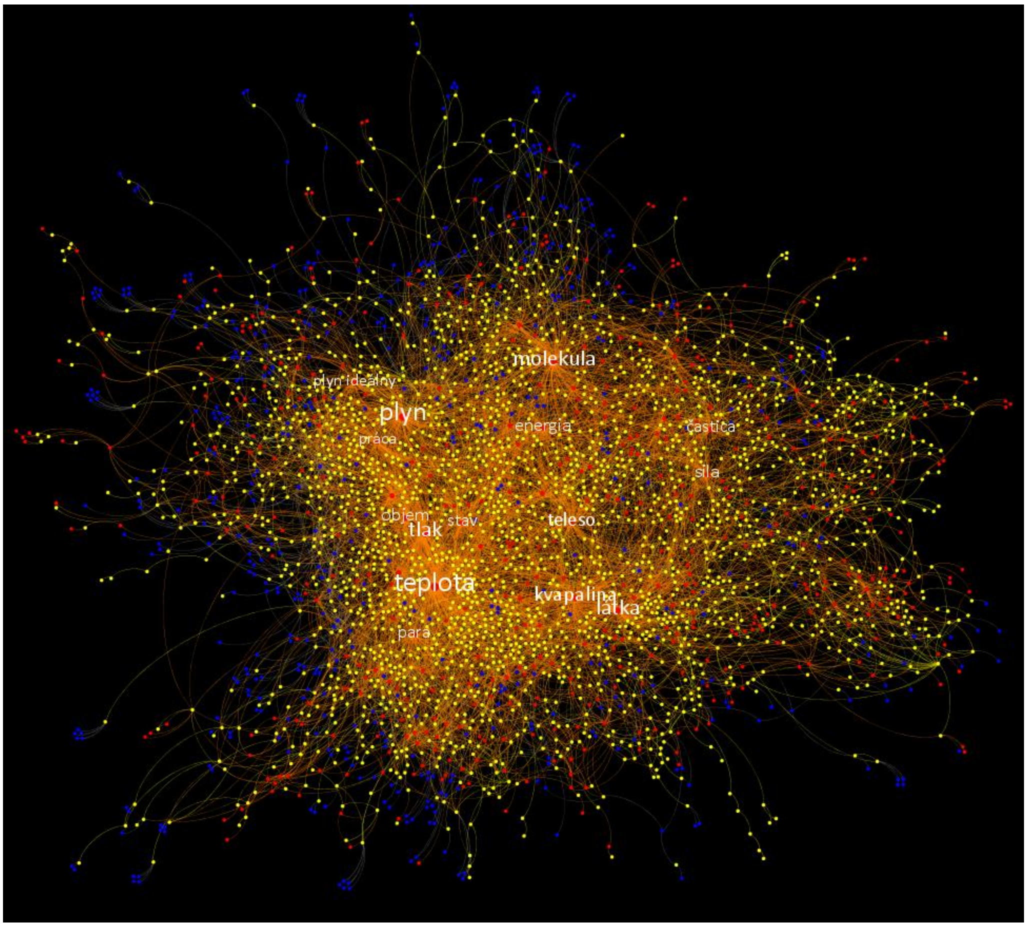

Figure 1. Resulting semantic network

The cluster analysis identified 13 clusters shown in Fig. 2. Three clusters (distinguished by yellow, purple and green color nodes in Fig. 2) were significantly larger the remaining ones. We investigated the properties of the whole tripartite network and the above mentioned three largest clusters. These properties are shown in Table 1.

Table 1: Network properties and individual clusters, where N – number of nodes, k – number of links, N_1 – number sentences , N_2 – number terms, N_3 – number of formulae, $\langle k \rangle$ – average number of links, d – diameter, $\langle d \rangle$ - average length, α – parameter describing the network.

	N	k	N_1	N_2	N_3	$\langle k \rangle$	d	$\langle d \rangle$	α
resulting network	3271	9225	2265	479	527	5,64	13	4,492	2,324
yellow cluster	1438	3711	1018	209	211	5,161	12	4,319	2,331
purple cluster	1316	3200	913	168	235	4,863	13	4,267	2,364
green cluster	359	711	248	60	51	3,961	10	4,207	2,568

We confirm that many nodes in the network have one or two links, and there are also several nodes that have significantly more connections (these are in particularly terms) – it is typical for scale invariant distribution. More detailed analysis shown that the network and also its largest clusters have scale invariant distributions (for the parameter *α* see the last column in Table 1).
Physics terms occurring most frequently in the text are: teplota (temperature – with 395 links), plyn (gas – 354), molekula (molecule – 237), látka (substance – 233), tlak (pressure – 222), kvapalina (liquid– 201), teleso (body – 185), energia (energy – 157), objem (volume – 155), sila (force – 140).

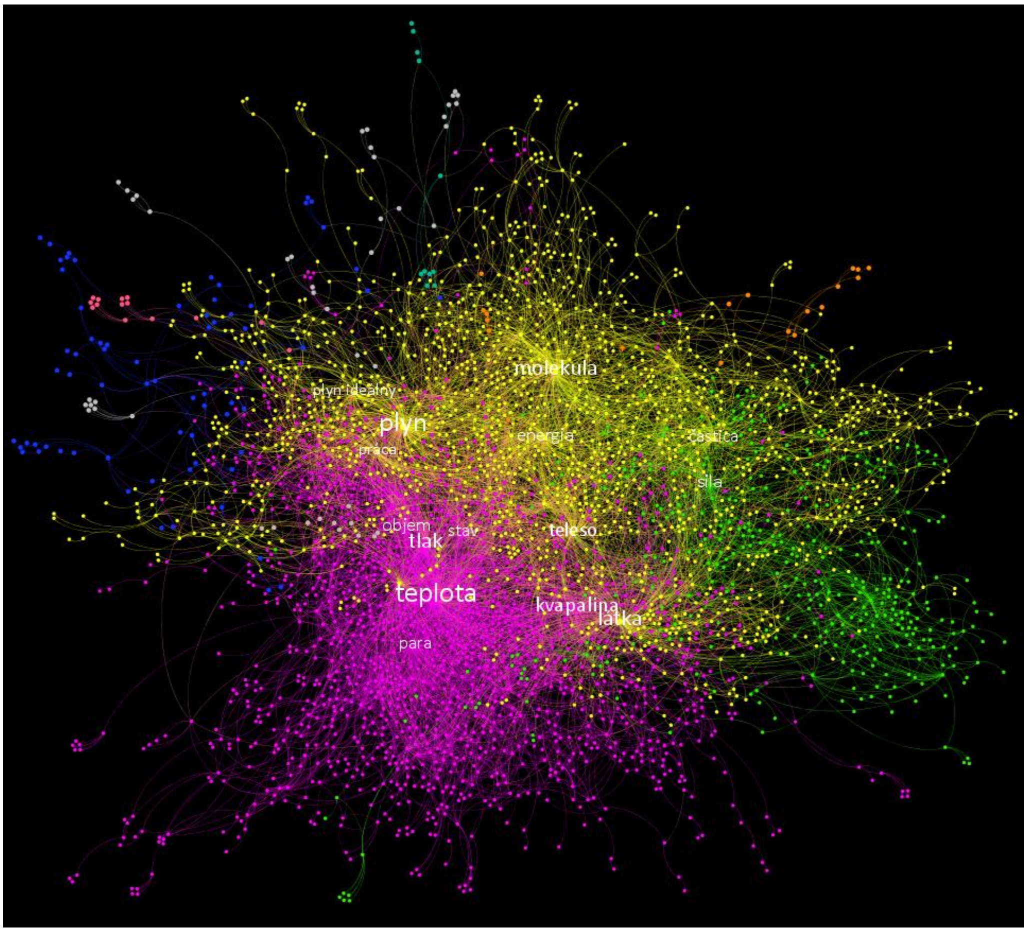

Figure 2. Semantic network divided into 13 clusters

Physics terms occurring most frequently in the largest cluster (yellow nods in Fig.2) are: plyn (gas – with 248 links), molekula (molecule – 196), energia (energy – 142), látka (substance – 141), častica (particle – 120), teleso (body – 118), práca (work – 86), ideálny plyn (ideal gas – 81), vnútorná energia (internal energy – 76), rýchlosť (velocity – 75). These terms are characteristic for the description of gaseous substances.

Physics terms occurring most frequently in the second largest cluster (purple nods) are: teplota (temperature – with 284 links), tlak (pressure – 160), kvapalina (liquid – 147), objem (volume – 120), para (vapour – 111), stav (state – 96), hmotnosť (mass – 69), rovnovážny stav (equilibrium state – 59), nasýtená para (saturated steam – 57), hustota (density – 40). This cluster is focused on liquid substances.

Terms occurring most frequently in the third largest cluster (green nods in Fig. 2) are: sila (force – with 74 links), deformácia (deformation – 54), dĺžka (length – 21), ťah (tension – 21), dislokácia (dislocation – 20), predĺženie (dilatation – 19), pružnosť (elasticity – 19), pevné teleso (rigid body– 14), deformácia ťahom (tensile deformation – 12), pevnosť (strength – 12). This cluster describes solid substances.

It is a surprise for us that the three largest clusters contain too many term type nodes with a single link (108).

The remaining clusters are shown in Fig. 3. These clusters contained 158 nodes and 179 edges, representing only 4.83 % of nodes from the overall network structure. The network represents 10 islands that cannot be interconnected and the value of parameter *α* is 3,115.

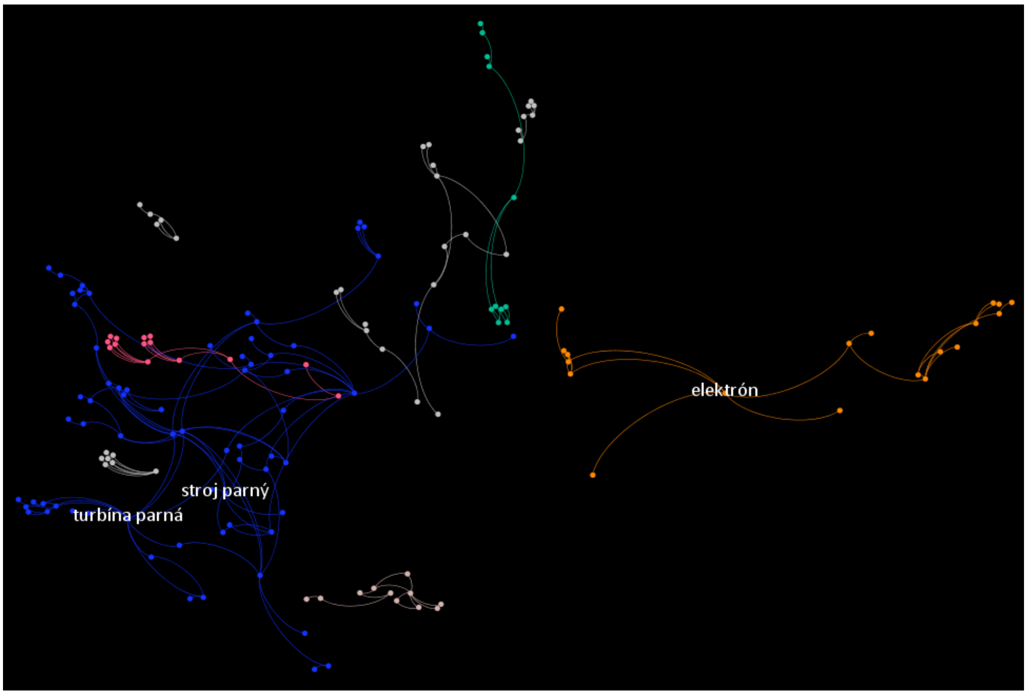

Figure 3. Remaining clusters

Terms occurring most frequently are: parná turbína (steam turbine – with 10 links), parný stroj (steam engine, as machine – 7), zážihový motor (ignition engine – 7), vývева (vacuum pump – 6), parný motor (steam engine – 6), reaktívny motor (reactive engine – 6), elektrón (electron – 5), plazma (plasma – 5), prúdový motor (jet engine – 5).

4. Discussion

The analyzed texts of the textbook [10] have a scale invariant distribution. Despite this fact, there are, in our opinion, too many nodes of the type term with only one connection in each network. Ideally, it should not be terms because the term (especially new terms and rarely used terms) gains its meaning in context only. If it is necessary to use a new term or a rarely used term in the textbook, it is better to introduce the term in an appropriate way (a nod with many connections). Terms with only one link are mostly terms that do not belong to the topic of the analyzed text.

In our analysis (provided by using Gephi) the resulting network consists of 13 clusters. The basic properties of the three largest clusters are shown in Table 1.

The three largest clusters represent the three basic state of matter - gaseous, liquid and solids state.

The remaining clusters represented only 4.83% of nodes from the resulting network. One of these clusters represented the fourth state of matter – plasma. Other clusters represented mostly examples and historical notes.

5. Conclusions

By analyzing the textbook, we visualized the curriculum of the textbook as a tripartite network.. In our opinion, from the viewpoint of the network distribution properties, the textbook and its largest parts (clusters) is written in appropriate manner for the student and facilitates him easier learning.
On the other hand, our analysis found too many term type nodes with only one link. The question arises whether these physical terms and names are important. The question is whether the textbooks are unnecessarily overwhelmed by physical terms and whether the names of persons do not burden the unnecessarily student.
We have analyzed the textbook for the sophomore class, so the nodes of type term with one link can be linked to a textbook from the previous year-class.
We can say that the network properties of the analyzed part meet the criteria of self-organized networks, despite the high occurrence of terms representing only one connection.
The presented method of analysis provides the possibility of analyzing the connection between physics textbooks for different year-classes and the connection of physics textbooks with the textbooks of related fields (mathematics, chemistry, biology, geography).
The number of usable analysis tools would be increased by processing a greater part of the teaching texts – a series of textbooks. We can analyze the robustness of the network (textbook), measured by their diameter or average distances between crucial parts of the network – one can simulate (randomly eliminating nods and edges) whether the textbook is fault tolerant.

Acknowledgments

This contribution was supported by UGA VII/10/2017.

References

[1] Albert, R.; Jeong, H.; Barabási and A. L. ”Error and attack tolerance of complex networks”, Nature 406, 2000, 378-482
[2] Barabási, A. L. and Albert, R. ”Emergence of scaling in random networks”, Science 286, 1999, 509-512
[3] Jeong, H. et.al ”The large-scale organization of metabolic network”, Nature 407, 2000, 651-654
[4] Barabási, A. L. ”Scale-free networks: A decade and beyond”, Science 325, 2009, 412-413
[5] Park, B. N. et.al ”Analysis of scientific item networks from science and biology textbooks”, Journal of Korea Contents Association 10, 2010, 427-435
[6] Čtrnáctová, H. ”Výběr a strukturace učiva chemie” [Selection and structuring of chemistry curriculum], Státni pedagogické nakladatelství, Praha, 1981, 176-188
[7] Barabási, A. L. ”Network Science: Graph Theory”, online: http://barabasi.com/f/625.pdf [accessed 21.11.2017]
[8] Barabási, A. L. ”Network Science: Evolving networks”, online: http://barabasi.com/f/621.pdf [accessed 21.11.2017]
[9] Teleki, A. ”Evolúciós kritériumok az oktatásban *A beiskolázás optimális feltételei*” [Evolution criteria in education Optimal conditions for school readiness], Pedagogická fakulta Univerzity J. Selyeho Komárne, 2015, 110-120
[10] Svoboda, E. et.al ”*Fyzika pre 2. ročník gymnázia*” [Physics for the 2nd year of high school], SPN, Bratislava, 1985
[11] Králiková, P. and Teleki, A. ”Physics textbooks from the viewpoint of network structures”, AIP Publishing, article number 040006, 2016, 1-7
[12] Čerňanský, P. et.al ”Fyzikálny slovník slovensko-anglický, anglicko-slovenský” [Physical dictionary Slovak-English, English-Slovak], Protonit s.r.o., 2007
[13] Garaj, J. et.al *”Fyzikálna terminológia”* [Physics terminology], SPN, Bratislava, 1987