THE CONCEPTUAL STRUCTURE OF PHYSICS TEXTBOOKS FOR SECONDARY SCHOOLS

Petra Králiková[1], Aba Teleki[2]

Abstract: Textbooks are an essential part of the learning process, therefore they need to be written in a way that is easy to understand. In real life, we often come across complex systems with scale invariant (power law) distributions, which display a surprising degree of tolerance against errors, i.e. degree of robustness. We are confident that knowledge organized in this manner is better for usage in textbooks and promotes easier learning as content would be more intelligible. Initially, we talk about the evolution of some networks, and then we deal with the differences between Poisson and scale invariant distribution in real networks. In conclusion, we are looking for connection between scale invariant distribution and Zipf's law.

Introduction

The primary school, secondary school and university students are coming across written texts every day. Written text play a special role in the improvement of a student's skills. It is important in formation of personality and acquisition of new knowledge or skills. Textbooks are essential part in the acquisition of knowledge, so they should be written intelligibly. We distinguish between intelligibility and readability (however, intelligibility of the text depends on its readability). The readability is influenced by many factors like length of words or length of sentences, number of words in the sentence, and number of new terminologies. To review textbook readability several indices exist, such as Gunning Fog, Flesch-Kincaid, Coleman-Liau, and SMOG index. Intelligibility is given by the structure of the text, the level of sentences, how sentences are connected, and what is the structure of these connections.

Interpretation of the curriculum in the textbooks makes up sentences. The sentences follow up on each other through words or parts of sentences (transitional words and transitional phrases). These links between sentences create the interpretation of the text and the interpretation is reflected in the structure of sentences. Studies of self-organization network structures showed, that they are governed by universal regularities. It is becoming apparent that the universal principle is the robustness of the network or the tolerance of the network against failures. We assume that the physics textbooks are more natural for students if they fulfill this universal principle. We are convinced that if the curriculum has scale invariant net structure, maximal learning capability can be achieved.

The evolution of the network

We can characterize the network as nodes (or vertices) which are linked to the edges (or links) as in Figure 1. There are many kinds of networks in real life:

- social networks – nodes are individuals and edges are the relationships between them,
- business networks – nodes are companies and edges are their business relationships,
- the internet – nodes are computers and edges are cables/wires connecting them,
- Word Wide Web or WWW – nodes are home pages and edges are link labels to other pages,
- biological networks (topological map of proteins) – nodes are proteins and edges are their direct cooperation in a living organism.

In our work we will study Slovak textbooks of physics, supposed to have been written more intelligibly. Nodes are sentences in the textbook and edges are transitional words and transitional phrases.

[1] Petra Králiková, Department of Physics, Constantine the Philosopher University, Tr. A. Hlinku 1, 94974 Nitra, Slovak Republic, petra.kralikova@ukf.sk

[2] Aba Teleki, Department of Physics, Constantine the Philosopher University, Tr. A. Hlinku 1, 94974 Nitra, Slovak Republic, ateleki@ukf.sk

Figure 1: Simple network with 8 nodes (or vertices) and 10 edges (links)

Source: Newman (2003)
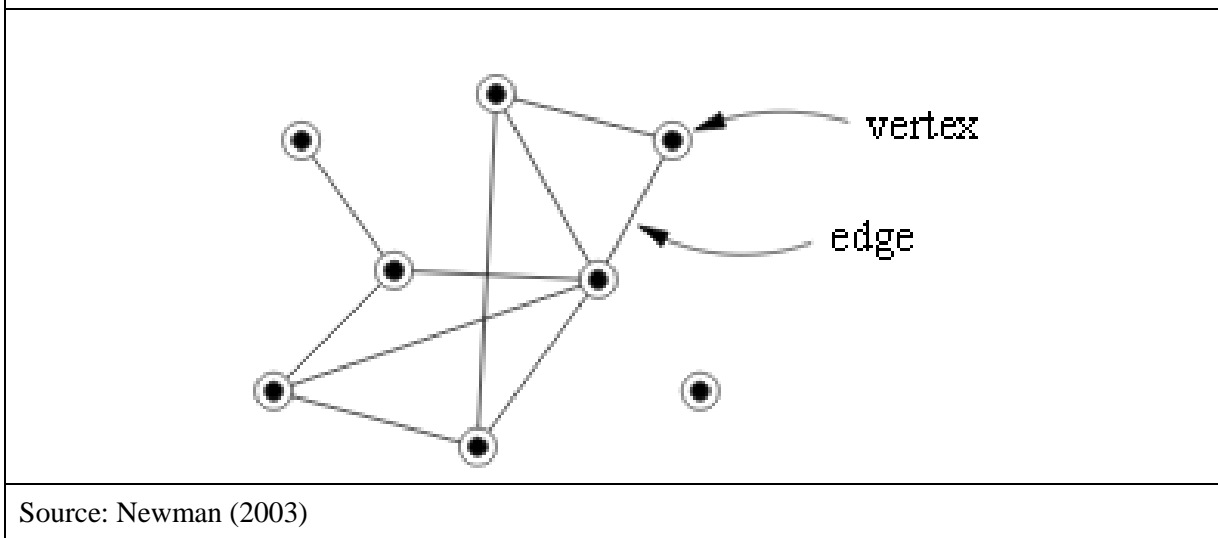
Barabási, Albert, & Jeong (2000) showed similarity of scale invariant networks for proteins and for the internet. In Figure 2a, we can see the topological map of proteins after approximately 1 billion years of development and in Figure 2b, we can see the topological map of the internet after 35 years of development.

Figure 2: Topological map of the proteins and of the internet

Source: Barabási et al. (2000)
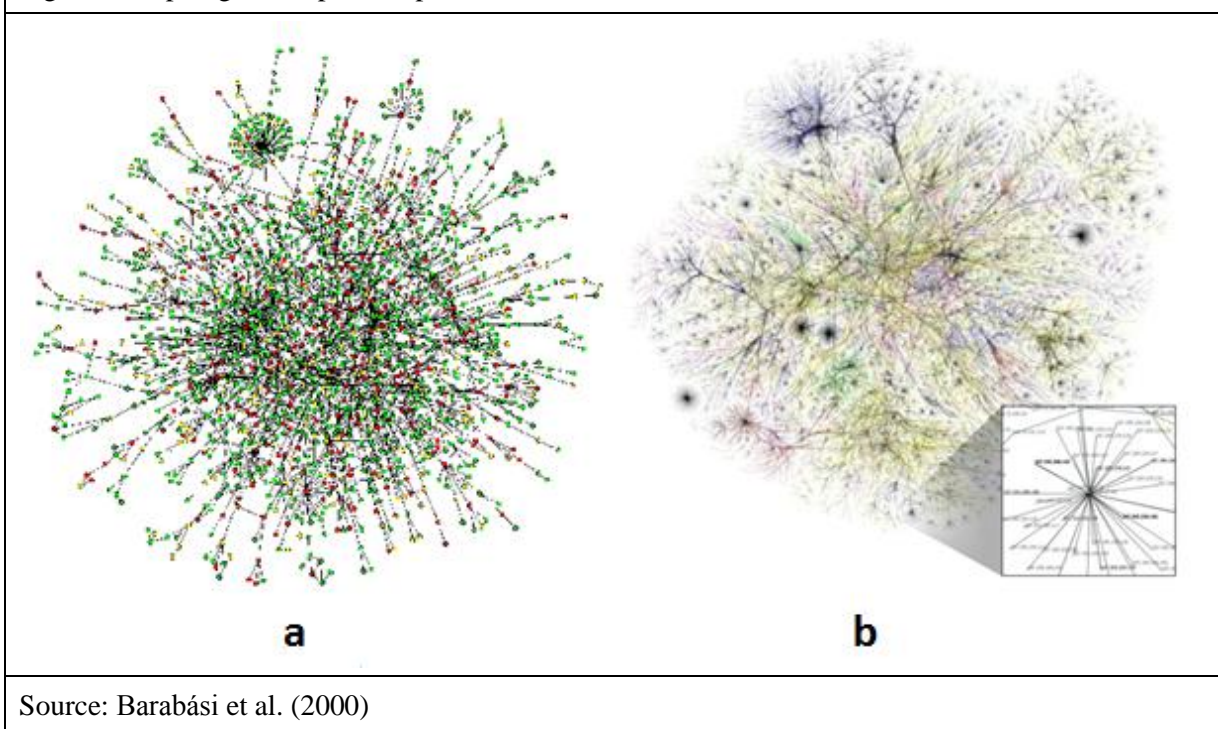
There are more models for self-organized networks. A well-known model is the Erdős-Rényi. In this model, number of links in one node has Poisson distribution as in Figure 3a. Most of the nodes have approximately equal number of links.

Barabási (2009) states that there is ever increasing evidence that this model is not satisfactory for real networks. In real networks, we observe power law distribution as in Figure 3b. In this Figure, we can see that most of the nodes have one or two links (black circles), but a few nodes have a large number of links (blue circle).

Figure 3: Poisson and power law distribution
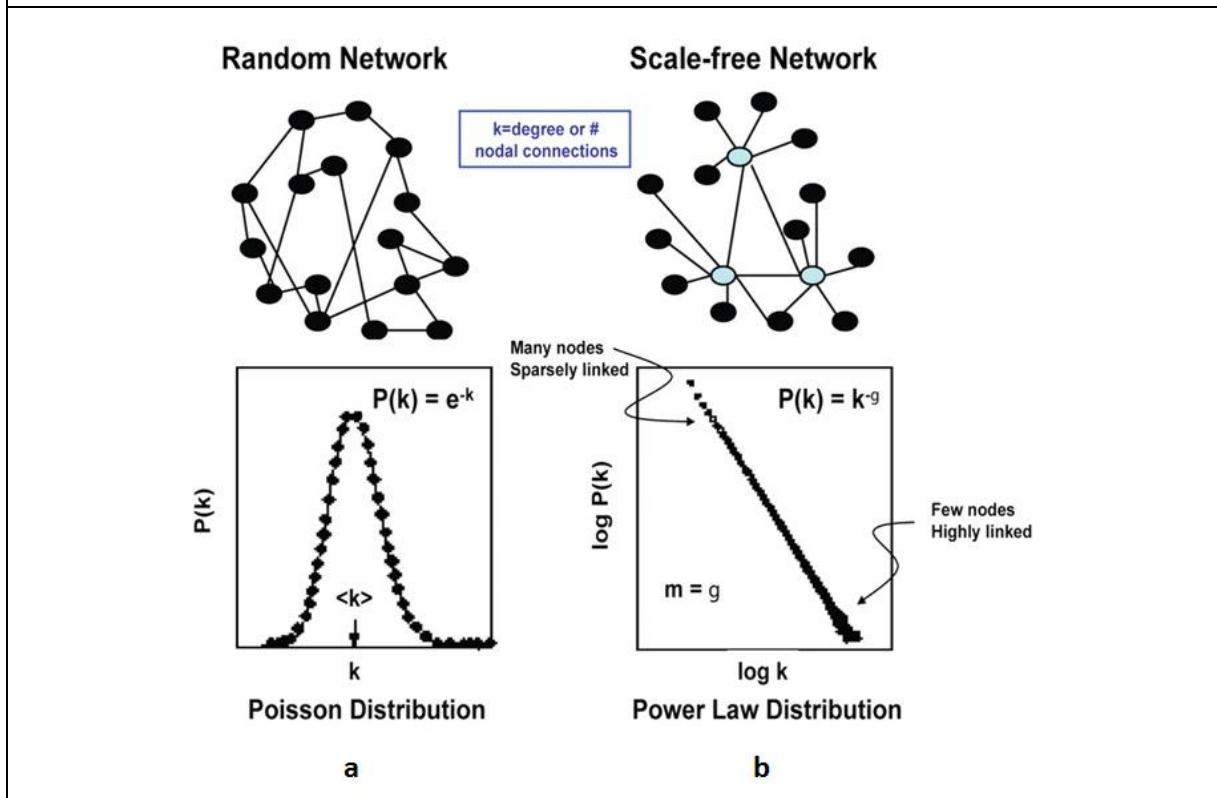
Source: Chan & Loscalzo (2012)

Barabási et al. (2000) and Albert, Jeong, & Barabási (1999) mapped the internet as a real network and proved that the World-Wide Web (WWW), the Internet and other large networks indicate that many systems belong to a class of inhomogeneous networks called scale-free networks or networks with power-law distribution. Barabási's collective analyzed other networks and showed that the most frequently self-organized networks realize scale invariant distribution too. They showed that these networks are resistant to spontaneous attack or failures and their capacity of extension is high.

We did illustrate the differences between the Poisson and power law distribution in Figure 4. Albert e al. (2000), Barabási et al. (2000), and Jeong, Tombor, Albert, Oltvai, and Barabási (2000) generated different kind of networks with the same number of nodes (10,000 nodes) and the same number of links (20,000 links). Barabasi investigated the influence of attacks and they measured the interconnectedness of networks under random attacks. The interconnectedness of a network is described by its diameter d, defined as the average length of the shortest paths between any two nodes in the network. The diameter characterizes the ability of two nodes to communicate with each other: smaller the value of d, shorter is the expected path between them. Networks with a very large number of nodes can have quite a small diameter; for example, the diameter of the WWW, with over 800 million nodes (Lawrence and Giles, 1999), is around 19 (Barabási & Albert, 1999), whereas social networks with over six billion individuals are believed to have a diameter of around Milgram (1967). The dependence is shown in Figure 4. Blue symbols characterize growth of the diameter of network under random attack, blue triangles describe Poisson distribution (random network) and blue squares describe power law distribution (scale invariant network). We can see that the interconnectedness decreases in random network, while the interconnectedness in scale invariant network almost doesn't change. Red symbols characterize directed (no random) attack, red diamonds describe random networks and red circles describe scale invariant network. We can say that scale invariant network is more sensitive on directed attack than on random network. Equal examples were modeled for N = 1000, N = 5000 and N = 20 000 nodes and they found equal values as in Figure 4. It means that results are independent of system size. Yin, Liu, Liu, and Li (2015) came to the same conclusion.

Figure 4: The influence of attack in random and scale invariant network
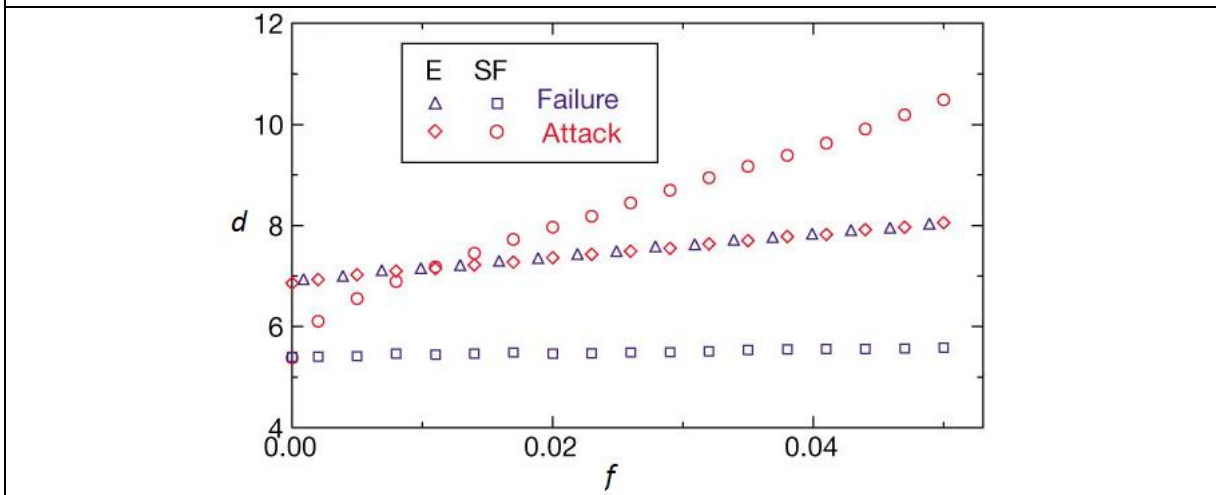
Source: Albert, Barabási, & Jeong (2000)

It has been shown, that many real networks have developed into scale invariant networks. We believe that if teaching material in textbooks (assigned for acquisition of knowledge) is organized as scale invariant networks, it will better activate the internal learning strategies of the brain, making the learning process easier.
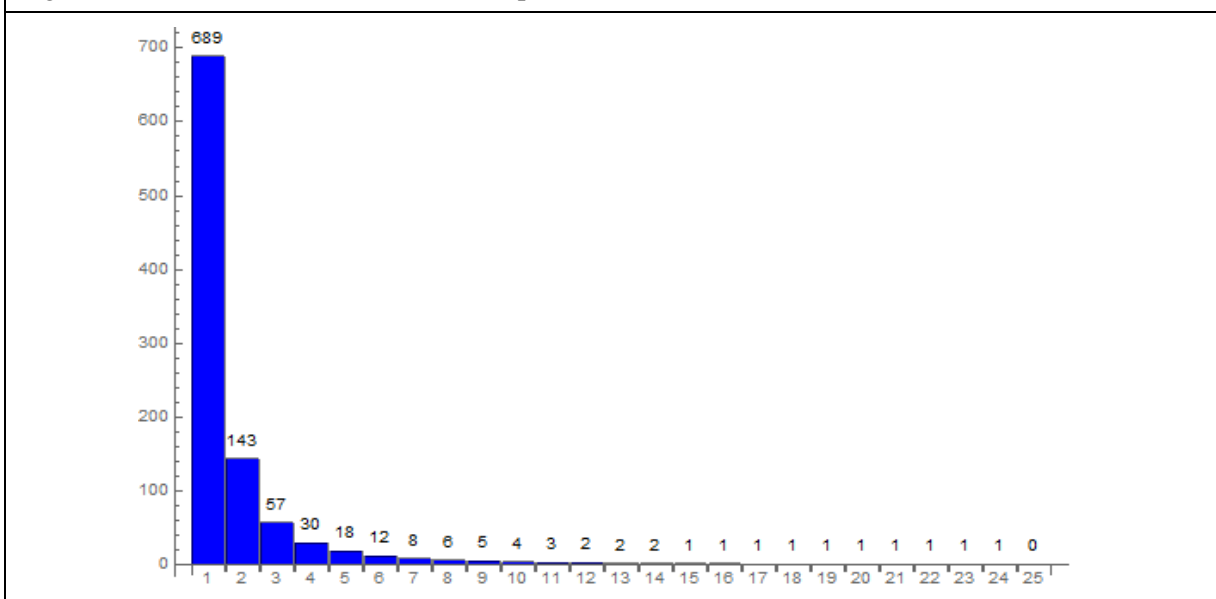
Scale invariant distribution and Zipf's law

The probability p in scale invariant (power law) distribution is given as

$$p \propto k^{-\alpha},$$

where k is the number of links to a node and parameter α characterizes the distribution. Yang and Zhao (2005) writes that the average value of the parameter α is usually between values $2 < \alpha < 3$, however, occasional exceptions exist. Teleki (2015) presented a histogram of links of particular distribution for power law distribution with number $N = 1,000$ nodes and parameter $\alpha = 2,2662$ as in Figure 5. Most of the nodes have one link and only one node has 24 links. Average number of links in one node is $k = 3$.

Figure 5: Scale invariant distribution with parameter $\alpha = 2,2662$

Source: Teleki (2015)

Barabási et al. (2000) in his research found that the internet and the web show scale invariant distribution. They investigated parameter α for "web, out" (outgoing links) and for "web, in" (incoming links), for the internet.

$$\alpha_{internet} = 2,48$$
$$\alpha_{web,out} = 2,45$$
$$\alpha_{web,in} = 2,1.$$

From the results we can see a very similar value of the parameter α. In all cases, the parameter α is in the range of $2 < \alpha < 3$. If the parameter α is different for two scale invariant networks they will be also visually different and this difference we can find out by mathematical analysis. If the parameter α is equal or similar for two scale invariant networks they will visually be very similar, although they may be different in details.

The fact that the parameter α takes the value of the above interval is also reflected in the learning process. Experiences show that a good teacher explains new concepts by using familiar concepts usually in two or three forms.

Now we describe Zipfˇs law and we seek connection between scale invariant distribution and Zipf's law. Adamic & Huberman (2002) showed that Zipf's law and Internet have many of the same features. They found that Zipf's law occurs on the internet very often.

Zipf's law demonstrates a relationship between the frequency of occurrence of certain events (frequency of words) and its order (if we rank events according to the size of the frequency). We have a rule, that the most frequently occurring word in the text is occurring x times, the second most frequently occurring word in the text is occuring x/2 times, the third most frequently occurring word in the text is occurring x/3 times, and so on. This law was described by Horecký (1961) who said that on plotting the above mentioned values, we will get a hyperbola as shown in Figure 6a. The frequency of each word is inversely proportional to its rank in the frequency table. The product of the frequency and the rank is approximately constant

$$fr = k$$

where f is the frequency of the word in the text, r is the rank of the word, and k is a constant.

Zipf's law is named after G. K. Zipf, but J. Estoup and E. K. Condom had studied this issue for a long time. They investigated this dependence more precisely because they examined the angle of inclination γ in log-log scale as we can see in Figure 6b. The value γ is 1 if the angle is 45˚ as in Zipf's model.

Figure 6: Zipf's law

Source: Horecký (1961)
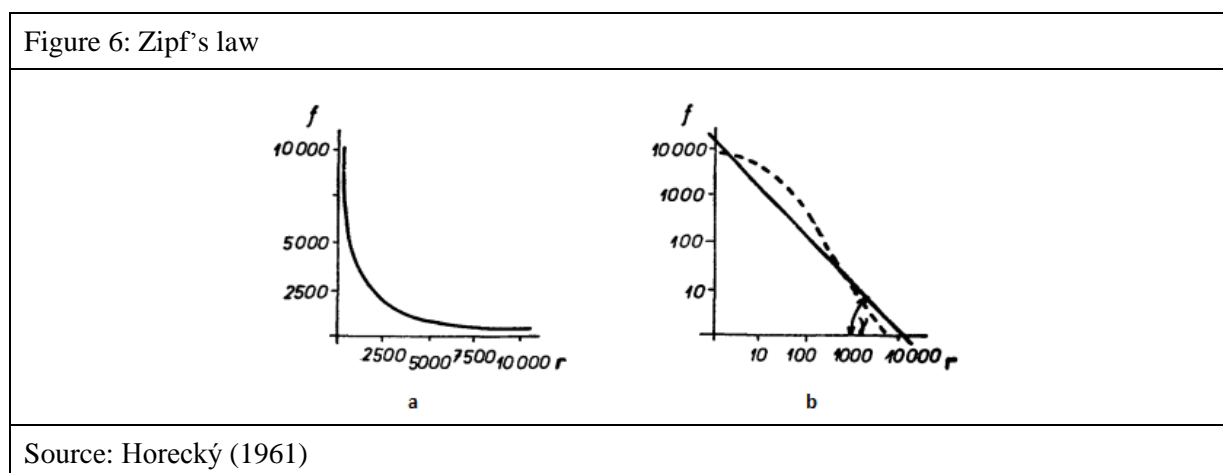
More accurate description of the frequency and rank is $f\, r^{\gamma} = $ const. or

$$f_r = \frac{k}{Nr^{\gamma}}$$

If we want relative values, we have to divide this formula by the total number of words, N. The ratio of the number of favorable cases f_r to all possible cases N characterizes the probability p_r, such that

$$p_r = \frac{k}{r^\gamma}$$

The form $p_r = k. r^{-\gamma}$ is very similar to the form for scale invariant distribution.

Results and Discussion

Several studies showed that real networks are mostly developed in scale invariant distribution. We can see this distribution in the Web, topological map of proteins and the Internet, therefore we expect it in physics textbooks for secondary schools as well. We will study every sentence, form, image, and graph in several Slovak physics textbooks and analyze in detail how they are interlinked by transitional words or transitional phrases. We intend to verify our assumption by close test.

Conclusion

In many real networks we can see the same universal regularity. This regularity shows that self-organization networks have scale invariant distribution. Complex systems which are scale invariant, have a surprising degree of tolerance to errors. We assume that the brain has evolved in the same way. In our opinion, if school work (intended for acquiring knowledge) has this structure, learning process will be easier. Detailed mapping of the conceptual structure of textbooks is useful because we want to find to what degree it supports the optimal strategy of learning. We want to find out if the detailed structure of concepts in the stylized text creates scale invariant network.

The goal of our next work is to make the networks analysis of physical textbooks for secondary schools using the above described criteria because the authors are not aware that such analysis on the Slovak physics textbooks was realized. In this respect, our country falls behind in research compared to the Anglo-Saxon countries, saying nothing about practical experience and development of automated tools (see project CoMetrix). These tools analyze the text and also offer alternatives to improve the quality of the analyzed texts (the teaching texts and the texts for the management companies). We must state that these tools do not use the principles revealed by Barabási's working group, therefore, we expect to contribute not only for authors of Slovak texts but also of texts in other languages.

Acknowledgement

References

Adamic, L. A., & Huberman, B. A. (2002). Zipf's law and the Internet. *Glottometrics 3*, 143-150.

Albert, R., Barabási, A. L., & Jeong, H. (2000). Error and attack tolerance of complex networks. *Nature 406*, 378-482.

Albert, R., Jeong, H., & Barabási, A. L. (1999). Diameter of the World-Wide Web. *Nature 401*, 130-131.

Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science 325*, 412-413.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science 286*, 509-512.

Barabási, A. L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: topology of the world wide web. *Physica A 281*, 69-77.

Chan, S. Y., & Loscalzo, J. (2012). The emerging paradigm of network medicine in the study of human disease. *Circulation Research*, 363-364.

Horecký, J. (1961). Využitie matematických metód v jazykovede [Use of mathematical methods in linquistics]. *SAV 5*, 257-269.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic network. *Nature 407*, 651-654.

Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature 400*, 107-109.

Milgram, S. (1967). The small-word problem. *Psychology Today 2*, 60-67.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review 45*, 167-256.

Teleki, A. (2015). A beiskolázás optimális feltételei [Optimal conditions for school readiness]. *J. Selye Univerzity in Komárno*, 110-120.

Yang, S. J, & Zhao, H. (2005). Generating multi-scaling networks with two types of vertices. *Physica A 370*, 863-868.

Yin, R. R., Liu, B., Liu, H. R., & Li, H. R. (2015). Research on invulnerability of the random scale-free network against cascading failure. *Physica A 444*, 458-465.