

Concept networking methods based on ChatGPT & Gephi

Dmitry Lande

National Technical University of Ukraine – Igor Sikorsky Kyiv Polytechnic Institute

Leonard Strashnoy

The University of California, Los Angeles (UCLA)

Annotation

This work is dedicated to describing the methodology of forming semantic network structures related to the main characters of a book and various concepts of the subject area. The methodology is based on the use of the chatGPT system, trained by a linguistic network, so no formulas or explicitly assigned calculations are used in the article. The methodology includes visualization of the formed networks. For this purpose, the Gephi (gephi.org) program is used. The CSV format is suitable for loading data into the Gephi environment, so all prompts to Chat GPT are accompanied by a data format requirement. The article shows the possibility of building concept networks based on the use of Chat GPT, solving tasks that were previously solved with the involvement of large temporal and human resources. It demonstrates how means of intelligent text analysis and network analysis are integrated, as well as their visualization. For the first time, the emulation of a group of experts is proposed by repeatedly building prompts to the ChatGPT system. As a result, a comprehensive methodology is proposed and tested, which can be applied for network analysis of literary works and building models of various subject areas.

Keywords:

Chat GPT, Concept networks, Graph visualization, Book network analysis, Domain model, Artificial experts, Information law.

Introduction

In today's world, the ability to manage and process this data has become a top priority for businesses and individuals alike. Big Data is revolutionizing our understanding of the world, but extracting meaningful information from such vast amounts of data is a worthy task. This is where ChatGPT comes in, an artificial intelligence system that allows users to efficiently manage and analyze their Big Data.

The ChatGPT system simplifies the processing and analysis of complex data, allowing users to quickly and efficiently gain the necessary knowledge, solving completely different problems.

The main objective of this work is to show the capabilities of ChatGPT for solving problems of semantic analysis and visualization, which allows it to be considered a useful tool for network analysis of documents, and the formation of domain models.

Network analysis and visualization are essential tools for understanding relationships between data points and uncovering hidden patterns that can help individuals and businesses to optimize

their workflow.

For network analysis and visualization, such powerful network analysis and visualization tools as Neo4j and Gephi can be used in conjunction with the ChatGPT system. Neo4j is a powerful server graph database that allows you to efficiently store and process large amounts of related data. More user-friendly, Gephi is an interactive personal platform that makes it easy to explore and visualize complex networks. It is this tool that will be used to demonstrate the examples in this article.

There are many works and concepts on the significance of such networks for science and practice, and special programs and languages for describing data have been created [1], [2]. The methodology allows the creation of networks, and graph structures with different topological features: simple undirected, weighted, directed weighted, and directed with labeled edges. And of course, the methodology is based on the use of the chat GPT system, trained on a linguistic network, so no formula or explicitly defined calculation will be used in the article. This is "A New Kind of Science", as Stephen Wolfram wrote when he was talking about cellular automata [3]. By the way, today he also writes a lot about the possibilities of chat GPT [4]. The methodology also includes visualization of the created networks. For this, the Gephi program (gephi.org) is used - the most popular program for visualizing graph structures with a free license [5]. The CSV format is quite suitable for loading data into the Gephi environment, so all requests to chat GPT will be accompanied by a format requirement.

Formation of networks of characters

So, in our plans, we have the description of procedures for forming networks with different properties. Let's move on to the description of tasks and solutions. We will demonstrate how character networks are formed using the example of analyzing Thor. It should be noted that not every book was covered by ChatGPT during its training. Obviously, the system "knows" a book of this scale. To build a network, it is sufficient to obtain a CSV file and load it into the Gephi program.

Forming a simple undirected network

(Forming a simple undirected network. Find (for example, 20) main pairs of characters in the text and describe their connections)

The ChatGPT system can help in obtaining the contents of such a file in the required format (fields corresponding to character names separated by semicolons). To do this, you can apply, for example, this prompt to the ChatGPT system:

Prompt 1	□ Give me the 20 most common pairs of Torah (Bible) characters in the
-----------------	---

format "character 1; character 2". Each entry on a new line

As a result, ChatGPT provides the following response:

- 1. Adam; Eve
- 2. Abraham; Isaac
- 3. Jacob; Esau
- 4. Jacob; Rachel
- 5. Jacob; Leah
- 6. Joseph; Potiphar's wife
- 7. Joseph; Pharaoh
- 8. Moses; Aaron
- 9. Moses; Pharaoh
- 10. Joshua; Caleb
- ...

After uploading the data into the Gephi system, we choose the size of nodes proportional to their degree (number of adjacent links) and divide the network into clusters based on modularity criteria, resulting in a visual graph (Figure 1).

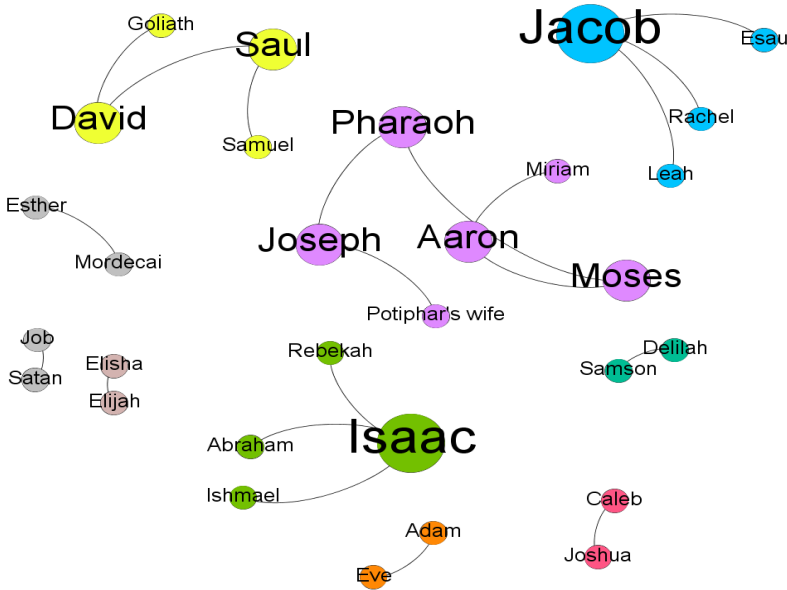


Figure 1. Undirected network of 20 main characters of Torah (Bible) according to ChatGPT program evaluation.

So, the first simple network has been built and it can be seen that the story of Passover occupies the central part of the Torah (Bible), which generally confirms the correctness of the selection of

Torah characters by the ChatGPT system.

Creating a weighted network

An already solved problem may be justifiably complicated by the requirement to indicate the weight of individual connections in the constructed network. Note that the "weight" is determined by the ChatGPT system according to its internal rules that are unknown to us. We can request the frequency of joint mentions of pairs of characters, but in this case, ChatGPT will most likely provide us with a program, most likely in Python, that can be used to calculate the frequencies. In our simplest case, we will rely on the artificial intelligence system and make the request.

Prompt 2	□ Give me the 20 most common pairs of actors (not letters) of the Torah (Bible) with the weight of the joint appearance of such pairs in the Torah in the format "actor 1; actor 2; weight". Each entry is on a new line.
-----------------	---

As a result, we will get an answer containing the requested information.

<p>As I mentioned earlier, the Torah (Bible) does not contain actors in the traditional sense, but it does contain a cast of characters such as prophets, leaders, and patriarchs. Here are the 20 most common pairs of characters in the Torah based on their joint appearance, along with their weights:</p> <ol style="list-style-type: none">1. Moses; God; weight: 7672. Abraham; God; weight: 2573. Aaron; Moses; weight: 1774. Isaac; Abraham; weight: 955. Jacob; Isaac; weight: 836. Joseph; Jacob; weight: 727. Sarah; Abraham; weight: 588. Joshua; Moses; weight: 559. Miriam; Aaron; weight: 4410. Pharaoh; Moses; weight: 41 <p>...</p>
--

To display a weighted network in Gephi, you can load the names as in the previous case, and the weight values in the "Edges" section of the "Data Laboratory" mode in this software.

Creating a directed weighted network

Considering the possibility of building a weighted network once again, let's further complicate the task by constructing a weighted directed network. We will determine the direction of the connection as the relationship of control between the characters, namely, the first character gives commands and controls the second. To do this, the following query is created:

Prompt 3	<p>□ Give me the 50 most common pairs of main characters (characters, not letters) of the Torah (Bible) with the weight of the joint appearance of such pairs in the Torah. Moreover, the first character commands something to the second. In the format "character 1; character 2; weight". Each entry is on a new line. I don't need the program code, but the result.</p>
-----------------	---

The ChatGPT system solves this problem in the following way:

<p>Here are the 50 most common pairs of main characters in the Torah (Bible) where the first character commands the second, along with their weight of joint appearance:</p> <ol style="list-style-type: none"> 1. Moses; Aaron; 214 2. God; Moses; 107 3. Moses; Joshua; 93 4. Moses; Pharaoh; 68 5. God; Aaron; 63 6. God; Joshua; 51 7. Moses; Bezalel; 50 8. Moses; Israelites; 45 9. Moses; Elders; 41 10. Moses; Levites; 36 <p>...</p>

Creating a signed network

In all previous examples, values were assigned to nodes, and the meaning of relationships was implied but not specified. Now we will show how to solve the problem of forming and displaying a network with edges that express various communicative relationships. To do this, we will create a prompt that explicitly asks the ChatGPT system to indicate the communication word between the characters. Naturally, a network with directed connections is topologically formed.

Prompt 4	<p>□ Give me the 10 most common pairs of main characters of the Torah. Moreover, the first character communicates something to the second.</p>
-----------------	--

	In the format "character 1; character 2; communication word". Each entry is on a new line. I don't need the program code, but the result.
--	---

As a result, we get the answer.

- | |
|---|
| <ol style="list-style-type: none"> 1. God; Adam; creation 2. God; Eve; command 3. Adam; Eve; temptation 4. Cain; Abel; jealousy 5. Noah; Ham; blessing 6. Abraham; Isaac; sacrifice 7. Isaac; Rebekah; marriage 8. Jacob; Esau; birthright 9. Joseph; his brothers; dreams 10. Moses; Aaron; leadership |
|---|

Let's ask to continue:

Prompt 4 (continuation)	<input type="checkbox"/> OK, another
-----------------------------------	--------------------------------------

We receive:

- | |
|---|
| <ol style="list-style-type: none"> 11. God; Noah; covenant 12. Abraham; Lot; separation 13. Jacob; Rachel; love 14. Jacob; Leah; deception 15. Joseph; Potiphar's wife; temptation 16. Moses; Pharaoh; Plagues 17. God; Moses; commandments 18. Aaron; Miriam; support 19. Isaac; Jacob; blessing 20. Hagar; Ishmael; exile |
|---|

To obtain weighted signed networks, one can use the following query:

Prompt 5	<input type="checkbox"/> List the 10 most important pairs of concepts in Torah only in the table format: "concept 1", "concept 2", and "importance". For every concept
-----------------	--

	<p>add inline in (Hebrew name, transliteration), Number rows. Add an inline English translation of concepts. For the "importance," column add (weight). Add another column that will represent the connection between pairs. Example: this row is related to rows x,y... and (explanation). Build a table that will represent a graph-like relationship.</p>
--	--

Emulating multiple experts through repeated prompt generation

(Another weighting method is just for directional networks)

The ChatGPT system can provide different answers when processing text at different times, all of which are reasonable from a human logic standpoint. Each of these answers can be perceived as the response of an artificial expert. By generalizing the answers of a multitude of such experts, a more complete and accurate answer can be obtained. It should be noted that sometimes the ChatGPT system may start giving the same answer for repeatedly entered queries, which should alert the user as this should not happen in such tasks. Presumably, the system has exhausted the resources allocated for the specific user. The solution to this problem is very simple - just reload the session. To demonstrate this possibility, a simple query can be repeated multiple times and the results can be combined.

<p>Prompt 6</p>	<p>□ Give me the 10 most common pairs of main characters of the Torah. Moreover, the first character commands (we can use the word communicate instead) something to the second. In the format "character 1; character 2". Each entry is on a new line. I don't need the program code, but the result.</p>
------------------------	--

Result of Chat GPT output:

<p>As an AI language model, I do not have access to information about character pairs in the Torah (Bible) where one character commands another. However, I can provide you with a list of the most common pairs of main characters in the Torah based on their appearances in the text. Please note that this list does not consider any commands or power dynamics between characters.</p> <ol style="list-style-type: none"> 1. Abraham; Sarah 2. Jacob; Esau 3. Jacob; Rachel 4. Jacob; Leah 5. Joseph; his brothers 6. Moses; Aaron 7. Moses; Pharaoh

- 8. Moses; Miriam
- 9. Joshua; Moses
- 10. Isaac; Rebekah

As a result of loading the merged CSV file into the Gephi environment, scaling node sizes, and clustering by modularity classes, we obtain a network of the main characters in the Torah (Bible) (Fig. 2).

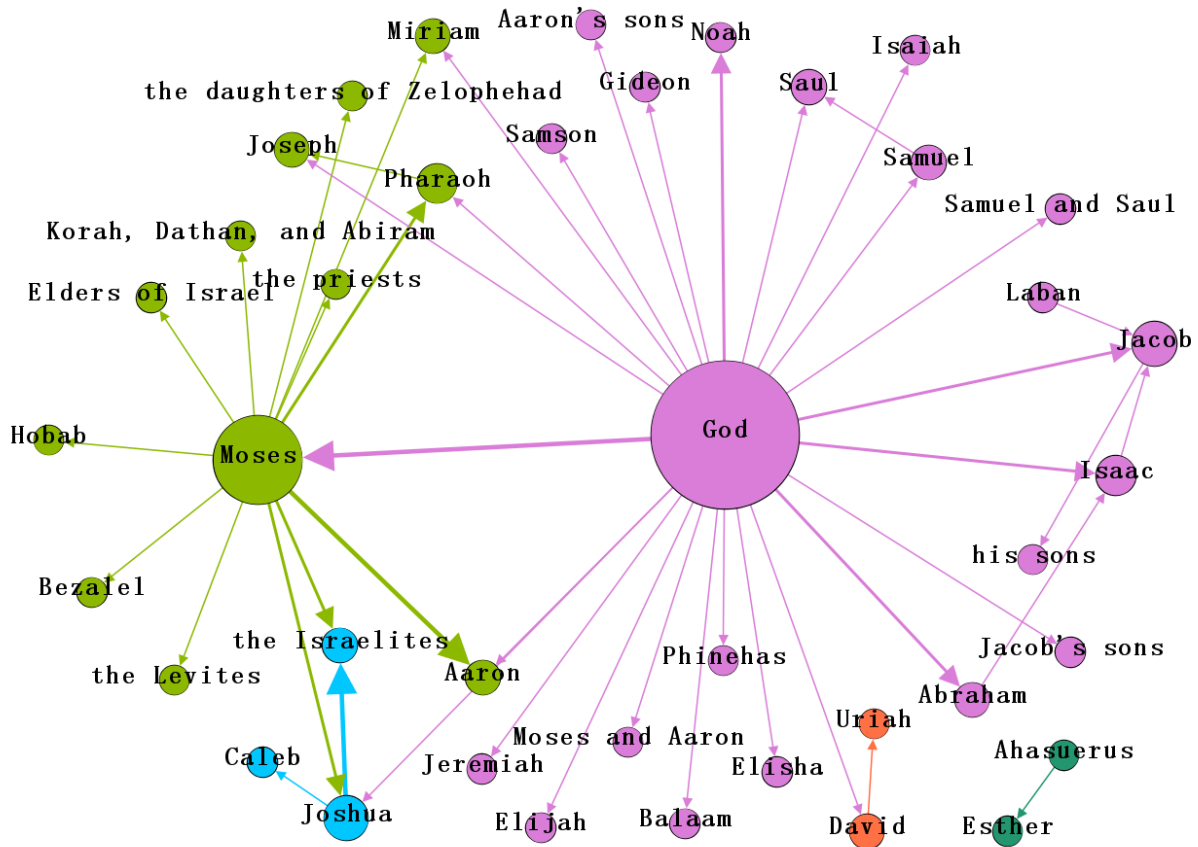


Figure 2. A network of main characters in the Torah is generated using the ChatGPT system by emulating an artificial expert swarm.

Formation of a network of concepts - Subject Domain model

Formation of a basic undirected network.

In all previous examples, we formed networks of concepts related to a specific book - the Torah (Bible), which was, of course, covered by the ChatGPT system during its training. Next, let's

consider the task of forming a network of concepts related to an entire subject area. Obviously, ChatGPT "knows" the basic concepts from the most important subject areas. For example, we are interested in the field of information law, so we will ask the ChatGPT system to find 40 pairs of the most popular concepts related to each other in this subject area. To do this, we will ask the ChatGPT system to perform the following query:

Prompt 7	□ Give me 40 pairs of the most important concepts of the subject area "Information law" in the format "concept 1; concept 2". Each entry is on a new line.
-----------------	--

As a result, we get:

<ol style="list-style-type: none">1. Data protection; Privacy2. Copyright; Fair Use3. Cybersecurity; Hacking4. Intellectual Property; Patents5. Freedom of Speech; Censorship6. Digital Rights Management; Encryption7. Trademark; Trade Secret8. Open Access; Public Domain9. Online Harassment; Cyberbullying10. Net Neutrality; Internet Service Providers <p>...</p>

Figure 4 shows the network formed after embedding the obtained data into the Gephi environment and subsequent aesthetic processing. Obviously, the formed network is weakly connected and incomplete. Let us assume that this network was obtained as a result of surveying only one artificial expert (as mentioned earlier).

Emulation of a set of experts by repeatedly generating prompts.

As previously shown in the analysis of Tor's characters, to ensure completeness of the set of concepts (or their pairs) and accuracy (insignificant connections will have little weight), one can use the possibility of launching a "swarm" of artificial agents. To do this, the previous prompt can be run several times and the answers aggregated into a CSV file. The network is expanded until it becomes sufficiently complete according to expert human assessments.

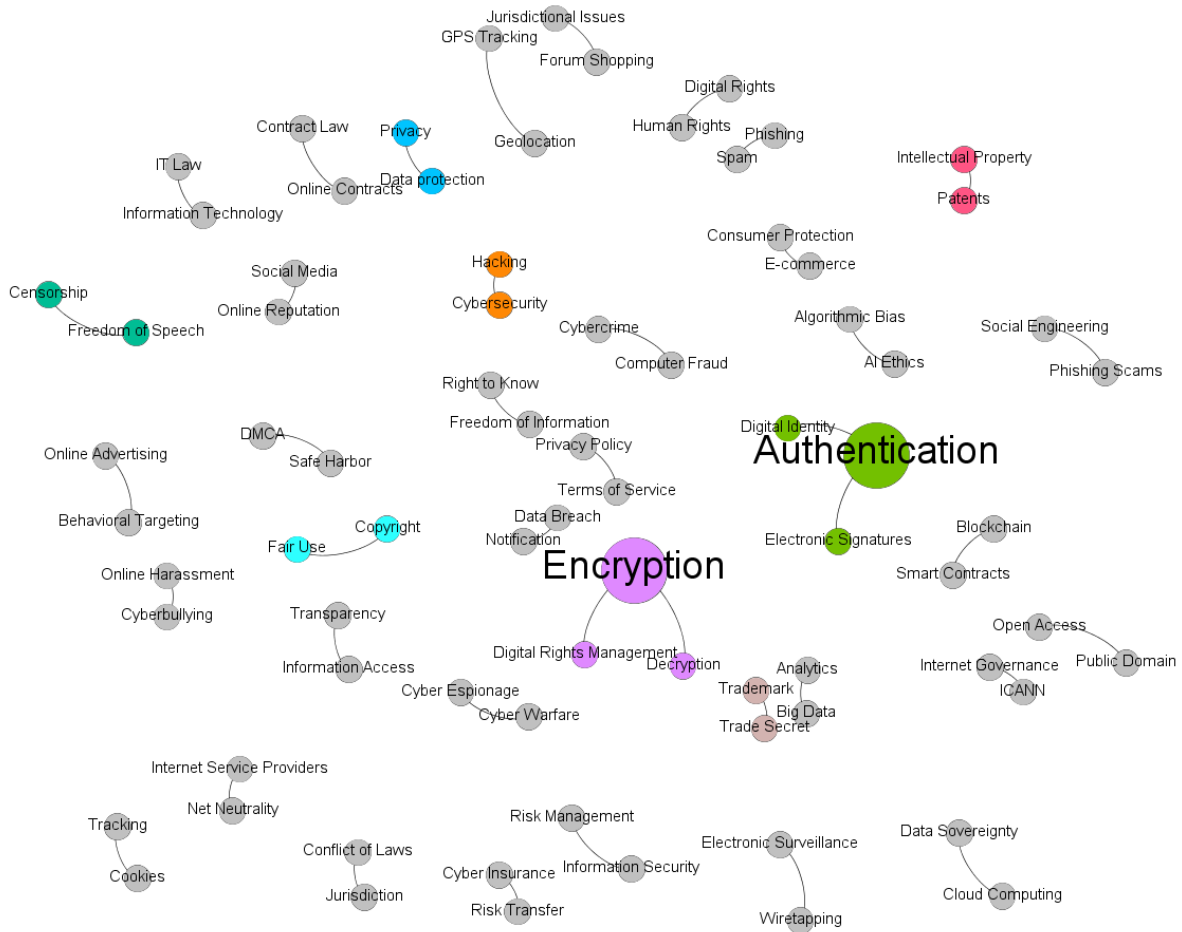


Figure 3. Primary network for the subject area "Information Law"

Conclusions

In this study, we have demonstrated

1. *Complex problems that previously required significant time and human resources can now be solved easily.*
2. *The convenience of using ChatGPT for building concept networks has been demonstrated, which has the potential to significantly reduce the resources that are usually needed to perform these tasks manually.*

Furthermore, we have shown how the integration of text analytics and network analysis tools can prove to be very useful in gaining insights from large amounts of unstructured data.

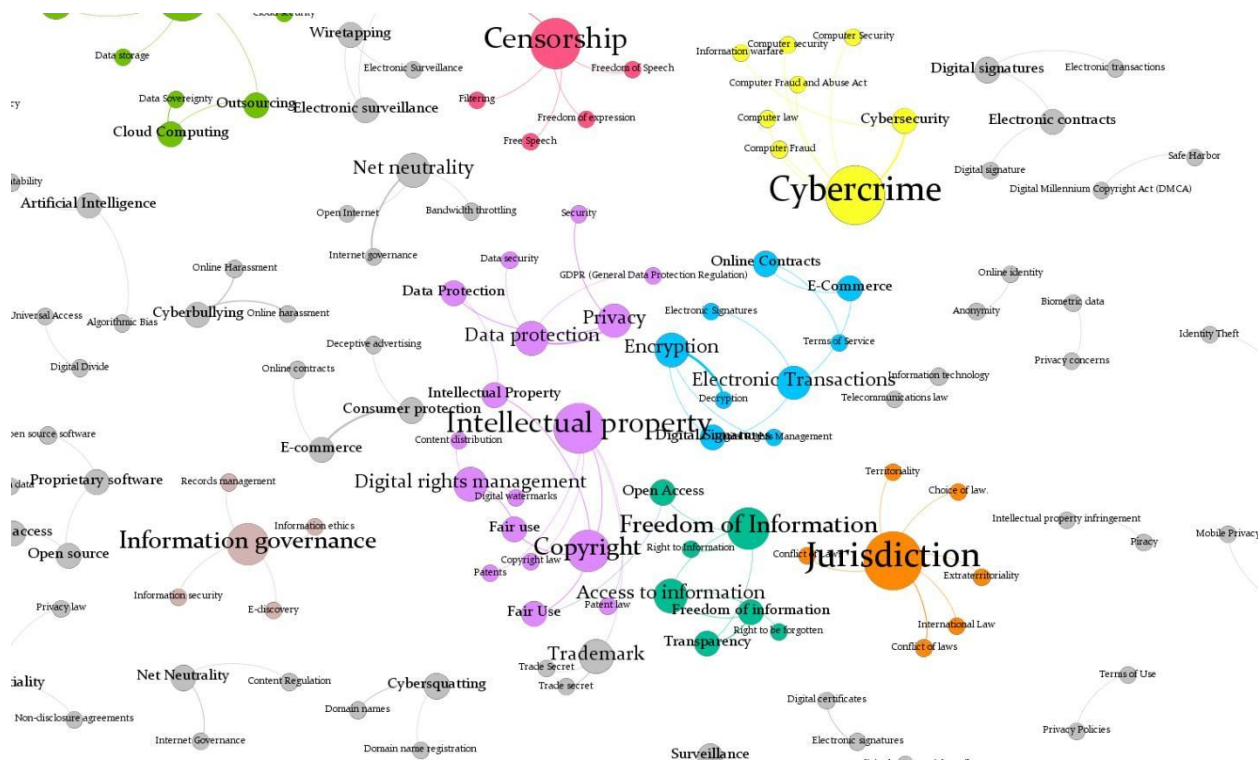


Figure 4. Visualization of a clustered network (using the example of the Information Law domain).

Figure 4 shows a central fragment of the network formed as a result of executing prompt 7 twenty times.

For the first time, an emulation of an "expert swarm" is proposed by repeatedly executing prompts to an artificial intelligence system.

One of the most interesting aspects of the study is the proposed emulation of groups of experts through the use of AI systems. This approach has the potential to greatly enhance the efficiency of knowledge extraction and to provide deeper insights into the structure and meaning of literary works across various subjects.

A methodology has been proposed that can be used for network analysis of documents and building models for various subject areas.

Overall, this study has demonstrated the practical applications of advanced AI technology in the field of text analysis and network visualization. It has shown how the use of machine learning algorithms can help us to unlock the previously hidden insights and patterns in textual data, and

to gain a deeper understanding of complex phenomena in different domains.

Limitations

However, it is important to note that this approach is not without its limitations. The interpretation of the results requires expertise in the subject area being studied, and there is still a need for human supervision to ensure the accuracy of the results.

Overall, it is clear that AI holds immense potential for transforming the field of text analysis and network visualization. While there are still many challenges that need to be addressed, this study represents an important step forward in unlocking the power of this technology to gain deeper insights into complex phenomena across a wide range of subjects.

Литература

1. T. Berners-Lee, J. Hendler, O. Lassila. The semantic web. Scientific American. Vol. 284, No. 5 (May 2001), pp. 34-43.
2. D.L. McGuinness, F. Van Harmelen. OWL web ontology language overview. W3C recommendation, 10 February 2004.
URL: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
3. St. Wolfram. "A New Kind of Science". – Wolfram Media, Inc. May 14, 2002. 1264 p.
4. St. Wolfram. "What Is ChatGPT Doing ... and Why Does it Work?". – Wolfram Media, Inc. March 9, 2023. 112 p.
5. Ken Cherven. "Mastering Gephi Network Visualization". – Packt Publishing, 2015. 378 p.
6. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K., "Asynchronous methods for deep reinforcement learning". In: International Conference on Machine Learning, 2016. pp. 1928–1937.
7. D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. "Mastering the game of go with deep neural networks and tree search". Nature, 529 (2016), p. 484.