ISSN: 1803-3814 (Printed), 2571-3701 (Online)
https://doi.org/10.13164/mendel.2020.1.015

ON VOYNICH ALPHABET ANALYSIS WITH RELATION TO THE OLD INDIAN DIALECTS

Ivan Zelinka[1,✉], **Tran Trong Dao**[2]

[1]Faculty of Electrical Engineering and Computer Science VŠB-TU, Department of Computer Science, Ostrava Poruba, Czech Republic
[2]Division of MERLIN, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh, Vietnam
ivan.zelinka@vsb.cz✉, trantrongdao@tdtu.edu.vn

Abstract

This paper is discussing our new research direction in the Voynich manuscript research. While our previous papers have been dealing with the research that has been based on fractal property analyses or graph properties analyses, where the graph has been constructed from the Voynich manuscript word sequences (Fig. 1), this paper discusses another kind of research on Voynich manuscript. This research is focused on the compassion of the letters or alphabets from Voynich manuscript with another selected alphabets from a different dialect, in that case, dialect from the Indian language. The reason is to point out the possibility that we can identify the origin of the Voynich manuscript alphabets based on the graphical conversion between letters from different dialects. Because this research is a very wide and deep topic, we publish in this paper only basic ideas, simulations and discuss all problems which have been found during those experimentation as well as outlining of the future directions of the research in an outlined way.

Received: 5 June 2020
Accepted: 10 August 2020
Published: 24 August 2020

Keywords: *Voynich manuscript, dialect clustering, image measure, similarity.*

1 Introduction

The Voynich manuscript is the most mysterious manuscript in the world. In fact, there is a 10 books (*Book of Soyga, Codex Seraphinianus, Hypnerotomachia Poliphili, the Oera Linda Book, the Ripley Scrolls, the Smithfield Decretals, the Rohonc Codex, the Red Book, Prodigiorum Ac Ostentorum Chronicon*) similar manuscript with Voynich manuscript in the first place. During its history or during in the existence of this manuscript there has been done a lot of research on his manuscript [3], [5],[6], [7], [8], [14], [11], [9] amongst the others, with the aim to translate this and understand, what kind of information is written inside. The people who tried to decode this manuscript were amateurs as well as professionals, including code-breakers from the second world war. No one succeeds so we can take into consideration a few hypotheses. The first one is that the manuscript is completely historical fake, which has been typical in the time of middle age when somebody wants to earn money by creating of false books and sell it for huge money to some rich people like the emperor Rudolf II was. Because everybody fails on the reading of this manuscript, it seems to be that this hypothesis cannot be true, however as some papers [1], used some statistical and natural language processing methods and has been shown that the properties of the Voynich manuscript seem to be a natural language. So probably the hypothesis dealing with the fake is not true. Another hypothesis is that the alphabets used in the Voynich manuscript manuscripts are artificially developed similarly like the Esperanto language. This is unlikely, as has been ruled by physical experimentation, this manuscript comes from 1404 - 1448, and it's unlikely that in that time somebody would be able to develop the full working artificial language, so we can't expect that this language will be completely unknown/artificial. The last hypothesis, we would like to discuss in this paper, is that this manuscript is written in the language, which has been retaken from some older dialects and adopted for European conditions. Of course, those things we are discussing here can be combined with some kind of primitive and encryption methods. However, because many researchers and the scientists equipped with very good computer power, they have been using, try to use different decryption methods on the manuscript, and never succeed. It is again unlikely, that this manuscript is encrypted and especially in the 15th century were encryption methods so weak, so if there would be any of those encryption methods applied, then definitely modern supercomputer would break it through. So the most likely, in our opinion, is the fact that Voynich manuscript is based alphabets, that are kind of mutation of alphabets from the previous languages and/or dialects. Simply our central hypothesis of this paper is that manuscript alphabets come from some old dialect and because dialect has been developing during the time including the shape of its alphabets it's highly likely to find

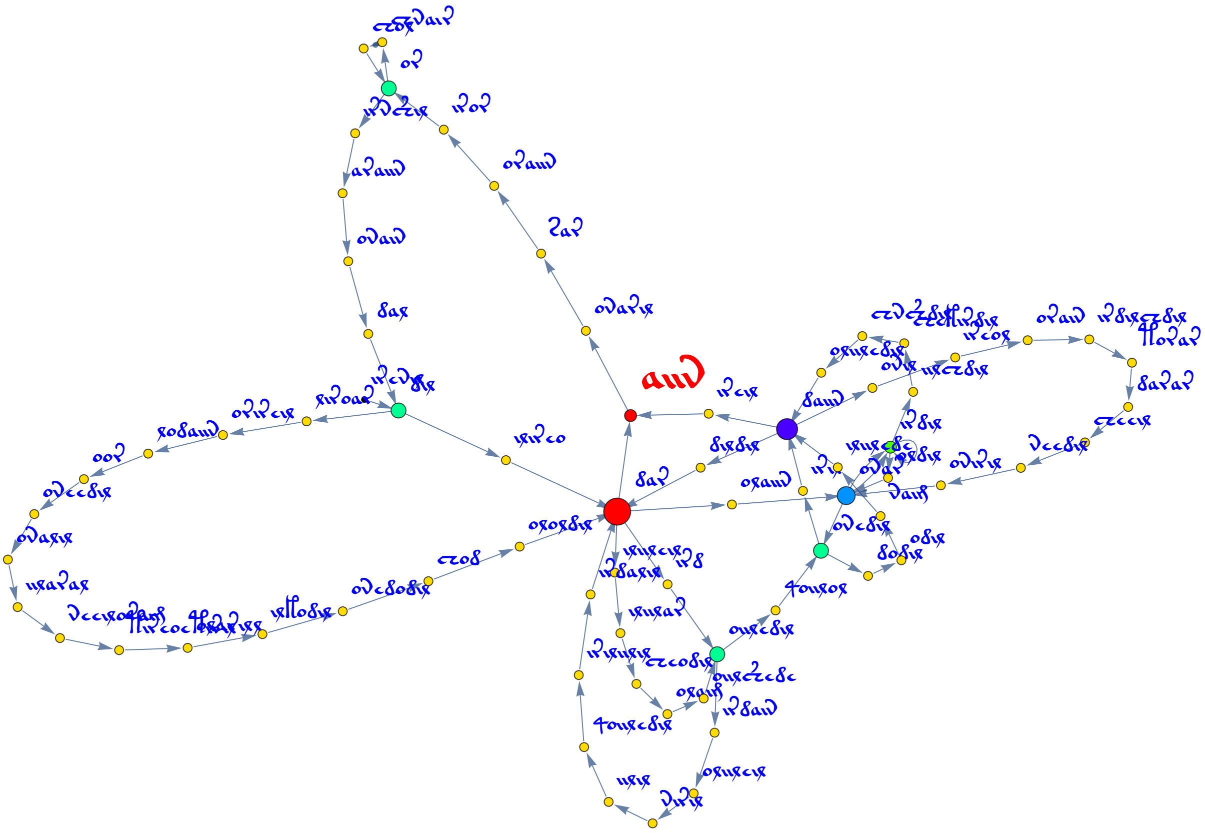
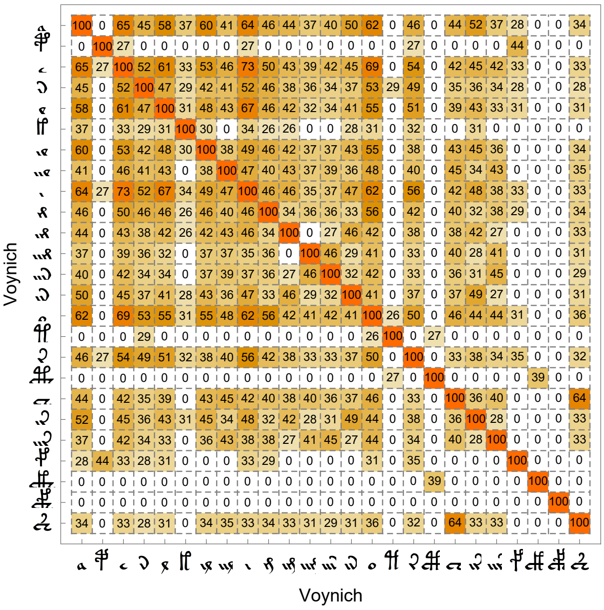


Figure 1: Network created from Voynich manuscript. A example of the small fraction.

the path of the mutation between different versions of giving dialect up to the modern version of them. This kind of researchers is actually under process in our research group [15] and in this paper we would like just to introduce some simple experiments which we used to graphically describe and analyse and compare alphabet from the Voynich manuscript with selected alphabets from other Indian dialects like Khoji for example. This paper is thus a kind of introduction into problems of the graphical comparison of the alphabets from different languages in order to find the most similar language based on the similarity measurement. The organisation of the paper is as follows. First, we briefly mention our previous research just by a few sentences and then we are continuing by the methods of the compassion of the letters as a graphical object to other dialects. The first experiments are based on the primitive compassion within the graphical objects, and then another more advanced are based on how to compare the graphical objects of their similarity by different measures (*EuclideanDistance, SquaredEuclideanDistance, NormalizedSquaredEuclideanDistance, ManhattanDistance, CosineDistance, CorrelationDistance, RootMeanSquare*) up to the most advanced method, which is based on the image feature extraction, regardless of object (i.e. letter) position in the figure. At the end in the conclusion, we discuss problems, identified during our experimentation, as well as the promising ways of the future research in that way.

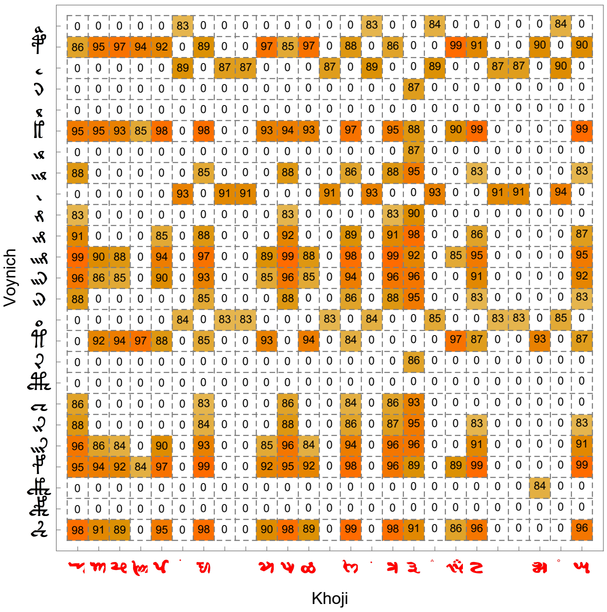
2 Motivation

The motivation for this research stems from the ever-existing fact that no one has yet deciphered, understood, or identified Voynia's manuscript in terms of language affiliation. It is therefore not known, if the manuscript is written in truly natural language, to which family of languages it belongs and thus what it could express, what culture it might reflect in its structure. Our goal now is not to decipher this manuscript, which, after all, no one has succeeded in doing so, but rather to point out the possibility of measuring the similarity of characters between individual languages and dialects. The aim is to identify possible connections between individual languages and dialects in the future to capture the gradual development. All this, we hope, could later help to identify Voynich manuscript in terms of language affiliation. The aim of this paper is to test a few basic methods for measuring the similarity of a language sign, pointing out which method is strong and which should be worked on further. Comparing the Voynich alphabet with the alphabets of other dialects, both medieval and ancient. It is a matter of working with linguistic texts, but in the form of graphical expressions, to which machine learning, classification and other methods are applied.



100	0	65	45	58	37	60	41	64	46	44	37	40	50	62	0	46	0	44	52	37	28	0	0	34
0	100	27	0	0	0	0	0	27	0	0	0	0	0	0	0	27	0	0	0	0	44	0	0	0
65	27	100	52	61	33	53	46	73	50	43	39	42	45	69	0	54	0	42	45	42	33	0	0	33
45	0	52	100	47	29	42	41	52	46	38	36	34	37	53	29	49	0	35	36	34	28	0	0	28
58	0	61	47	100	31	48	43	67	46	42	32	34	41	55	0	51	0	39	43	33	31	0	0	31
37	0	33	29	31	100	30	0	34	26	26	0	0	28	31	0	32	0	0	31	0	0	0	0	0
60	0	53	42	48	30	100	38	49	46	42	37	37	43	55	0	38	0	43	45	36	0	0	0	34
41	0	46	41	43	0	38	100	47	40	43	37	39	36	48	0	40	0	45	34	43	0	0	0	35
64	27	73	52	67	34	49	47	100	46	46	35	37	47	62	0	56	0	42	48	38	33	0	0	33
46	0	50	46	46	26	46	40	46	100	34	36	36	33	56	0	42	0	40	32	38	29	0	0	34
44	0	43	38	42	26	42	43	46	34	100	0	27	46	42	0	38	0	38	42	27	0	0	0	33
37	0	39	36	32	0	37	37	35	36	0	100	46	29	41	0	33	0	40	28	41	0	0	0	31
40	0	42	34	34	0	37	39	37	36	27	46	100	32	42	0	33	0	36	31	45	0	0	0	29
50	0	45	37	41	28	43	36	47	33	46	29	32	100	41	0	37	0	37	49	27	0	0	0	31
62	0	69	53	55	31	55	48	62	56	42	41	42	41	100	26	50	0	46	44	44	31	0	0	36
0	0	0	29	0	0	0	0	0	0	0	0	0	0	26	100	0	27	0	0	0	0	0	0	0
46	27	54	49	51	32	38	40	56	42	38	33	33	37	50	0	100	0	33	38	34	35	0	0	32
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	100	0	0	0	0	39	0	0
44	0	42	35	39	0	43	45	42	40	38	40	36	37	46	0	33	0	100	36	40	0	0	0	64
52	0	45	36	43	31	45	34	48	32	42	28	31	49	44	0	38	0	36	100	28	0	0	0	33
37	0	42	34	33	0	36	43	38	38	27	41	45	27	44	0	34	0	40	28	100	0	0	0	33
28	44	33	28	31	0	0	0	33	29	0	0	0	0	31	0	35	0	0	0	0	100	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	0	100	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
34	0	33	28	31	0	34	35	33	34	33	31	29	31	36	0	32	0	64	33	33	0	0	0	100

Figure 2: Similarity between alphabets from Voynich based on cosine distance measure.



0	0	0	0	0	83	0	0	0	0	0	0	0	0	83	0	0	84	0	0	0	0	0	84	0
86	95	97	94	92	0	89	0	0	97	85	97	0	88	0	86	0	0	99	91	0	0	90	0	90
0	0	0	0	0	89	0	87	87	0	0	0	87	0	89	0	0	89	0	0	87	87	0	90	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
95	95	93	85	98	0	98	0	0	93	94	93	0	97	0	95	88	0	90	99	0	0	0	0	99
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	0	0	0	0	0	0	0	0
88	0	0	0	0	0	85	0	0	0	88	0	0	86	0	88	95	0	0	83	0	0	0	0	83
0	0	0	0	0	93	0	91	91	0	0	0	91	0	93	0	0	93	0	0	91	91	0	94	0
83	0	0	0	0	0	0	0	0	0	83	0	0	0	0	83	90	0	0	0	0	0	0	0	0
91	0	0	0	85	0	88	0	0	0	92	0	0	89	0	91	98	0	0	86	0	0	0	0	87
99	90	88	0	94	0	97	0	0	89	99	88	0	98	0	99	92	0	85	95	0	0	0	0	95
96	86	85	0	90	0	93	0	0	85	96	85	0	94	0	96	96	0	0	91	0	0	0	0	92
88	0	0	0	0	0	85	0	0	0	88	0	0	86	0	88	95	0	0	83	0	0	0	0	83
0	0	0	0	0	84	0	83	83	0	0	0	83	0	84	0	0	85	0	0	83	83	0	85	0
0	92	94	97	88	0	85	0	0	93	0	94	0	84	0	0	0	0	97	87	0	0	93	0	87
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
86	0	0	0	0	0	83	0	0	0	86	0	0	84	0	86	93	0	0	0	0	0	0	0	0
88	0	0	0	0	0	84	0	0	0	88	0	0	86	0	87	95	0	0	83	0	0	0	0	83
96	86	84	0	90	0	93	0	0	85	96	84	0	94	0	96	96	0	0	91	0	0	0	0	91
95	94	92	84	97	0	99	0	0	92	95	92	0	98	0	96	89	0	89	99	0	0	0	0	99
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98	91	89	0	95	0	98	0	0	90	98	89	0	99	0	98	91	0	86	96	0	0	0	0	96

Figure 3: Similarity between alphabets from Voynich and Khoji dialect, based on binary distance measure.

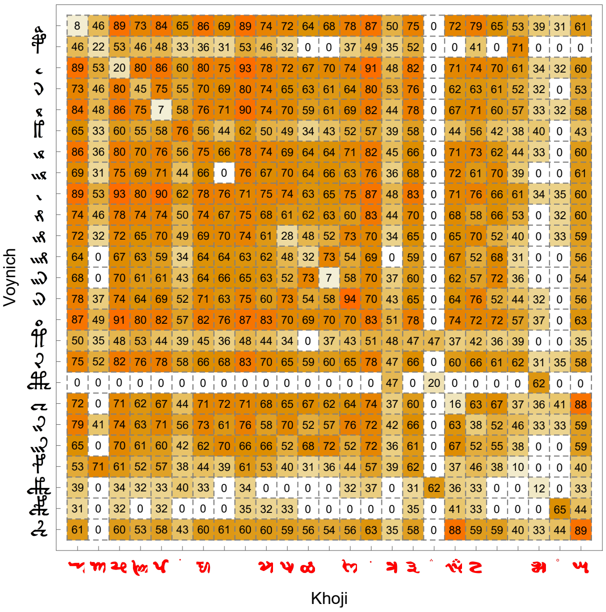
3 Experiment Design

Four of our experiments presented in our paper here, we have used standard computer equipment like PC with i7 processor, 8GB memory and *Mathematica* 11. Nothing else has been needed for our experimentation. We have tested different measurement methods for letter comparison as the *EuclideanDistance, SquaredEuclideanDistance, NormalizedSquaredEuclideanDistance, ManhattanDistance, CosineDistance, CorrelationDistance, RootMeanSquare.*

Several basic experiments were performed in this article. The first experiments were focused on processing the characters of the alphabet as graphic objects by reading them from their graphic form using a neural network and converting them into a matrix of ones and zeros. The black-and-white image thus obtained was then used to measure the distance between the individual characters of the alphabet. Several distance measurement methods were used, as mentioned above. These were applied first to the Voynich alphabet itself, then to the alphabet of Vounice and Khoji with different methods of measuring similarity. The second set of experiments was performed to describe the character of the Voynich alphabet not in the form of a binary matrix, but in the form of so-called image corresponding points, which are significant points that determine the character of a graphic object, as will be explained below. The description obtained in this way actually presented the given object in the form of a vector, and methods for measuring distances were again used for this description. It was tested on the Voynich alphabet to see if this method can classify letters that are related and similar. This was then also visualized in the form of a graph, where the morphology and gradual change of individual letters in terms of similarity can be clearly seen. In other words, these graphs express the visual similarity between the individual characters of the Voynich alphabet. In this experiment, we also visualized the alphabet in the form of a graph, where the letters are shown in the form of colored vertices. Color and size determine the meaning of a letter in a given alphabet in terms of measuring similarity. As will be explained below, the last set of experiments was again based on image corresponding points, and here we used basic hierarchical clustering methods to determine how some letters of the Voynich alphabet and the Indian Khoji dialect are graphically similar, as demonstrated in the figures at the end of the article. Thus, 3 basic sets of experiments were performed, which are described below.

4 Results

In the first experiments, it was necessary to convert the characters of Voynich alphabet into a graphical form as well as the alphabet of other dialects, in this case, Khoji and then convert these images back to a suitable description for computer processing. This is because we assume that not all dialects of the future will generate themselves. Still, we get their photos, so we tried to simulate the conditions that we got images of an element of the dialect from somewhere. In this case, it was necessary to process the images so that for example Khoji generated by Mathematica software had to be saved as a binary image format loaded into mathematics and converted into a matrix that described the object in this image as the matrix of 1 and 0. We followed the description in the form of zeros and ones and to convert images into such form a neural network was used. The image was to the grey degrees, which were then adjusted to black and white. Thus modi-



8	46	89	73	84	65	86	69	89	74	72	64	68	78	87	50	75	0	72	79	65	53	39	31	61
46	22	53	46	48	33	36	31	53	46	32	0	0	37	49	35	52	0	0	41	0	71	0	0	0
89	53	20	80	86	60	80	75	93	78	72	67	70	74	91	48	82	0	71	74	70	61	34	32	60
73	46	80	45	75	55	70	69	80	74	65	63	61	64	80	53	76	0	62	63	61	52	32	0	53
84	48	86	75	7	58	76	71	90	74	70	59	61	69	82	44	78	0	67	71	60	57	33	32	58
65	33	60	55	58	76	56	44	62	50	49	34	43	52	57	39	58	0	44	56	42	38	40	0	43
86	36	80	70	76	56	75	66	78	74	69	64	64	71	82	45	66	0	71	73	62	44	33	0	60
69	31	75	69	71	44	66	0	76	67	70	64	66	63	76	36	68	0	72	61	70	39	0	0	61
89	53	93	80	90	62	78	76	71	75	74	63	65	75	87	48	83	0	71	76	66	61	34	35	60
74	46	78	74	74	50	74	67	75	68	61	62	63	60	83	44	70	0	68	58	66	53	0	32	60
72	32	72	65	70	49	69	70	74	61	28	48	52	73	70	34	65	0	65	70	52	40	0	33	59
64	0	67	63	59	34	64	64	63	62	48	32	73	54	69	0	59	0	67	52	68	31	0	0	56
68	0	70	61	61	43	64	66	65	63	52	73	7	58	70	37	60	0	62	57	72	36	0	0	54
78	37	74	64	69	52	71	63	75	60	73	54	58	94	70	43	65	0	64	76	52	44	32	0	56
87	49	91	80	82	57	82	76	87	83	70	69	70	70	83	51	78	0	74	72	72	57	37	0	63
50	35	48	53	44	39	45	36	48	44	34	0	37	43	51	48	47	47	37	42	36	39	0	0	35
75	52	82	76	78	58	66	68	83	70	65	59	60	65	78	47	66	0	60	66	61	62	31	35	58
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	20	0	0	0	0	62	0	0
72	0	71	62	67	44	71	72	71	68	65	67	62	64	74	37	60	0	16	63	67	37	36	41	88
79	41	74	63	71	56	73	61	76	58	70	52	57	76	72	42	66	0	63	38	52	46	33	33	59
65	0	70	61	60	42	62	70	66	66	52	68	72	52	72	36	61	0	67	52	55	38	0	0	59
53	71	61	52	57	38	44	39	61	53	40	31	36	44	57	39	62	0	37	46	38	10	0	0	40
39	0	34	32	33	40	33	0	34	0	0	0	0	32	37	0	31	62	36	33	0	0	12	0	33
31	0	32	0	32	0	0	0	35	32	33	0	0	0	0	0	35	0	41	33	0	0	0	65	44
61	0	60	53	58	43	60	61	60	60	59	56	54	56	63	35	58	0	88	59	59	40	33	44	89

Figure 4: Similarity between alphabets from Voynich and Khoji dialect, based on cosine distance measure.

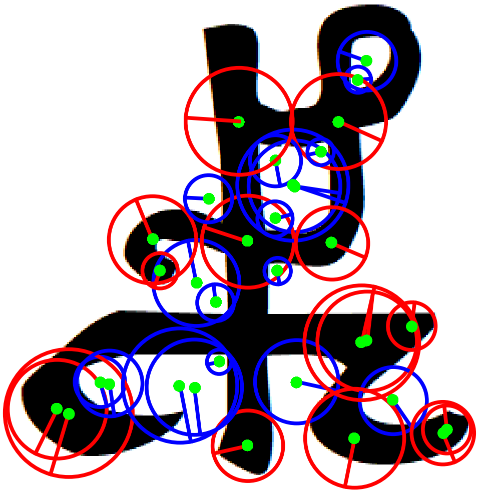


Figure 5: The *image corresponding points* principle on selected Voynich alphabet.

fied images which were de facto in the form of matrices we then worked on, were used by functions of image processing, which are part of the software Mathematica which helped us to measure different types of distances and similarities between alphabets images or if you want, between the letter of the alphabets. Several experiments, with increasing complexity and sophisticated approach, has been done here. We have started with standard classical compassion of the letters being represented as a picture. In the first experiment, we have used Voynich letters as graphical objects. We have generated whole alphabets in Mathematica software and convert it into unified images of the black-white colour each of them as a separated file. That set of the Voynich alphabets and its similarity measure has been used to create Fig. 2. The similarity between alphabets in the Voynich manuscript the picture is visible and shows that the highest similarity 100% is only between letters which are the same another similar similarity between letters are different and lower of course. Those fields which contain 0 are under our threshold of similarity, manually adjusted. This has been done just to demonstrate whether in or not whether the method can recognize the similarity between some letters especially with the same letter and with what kind of success it is possible to see also the highest similarity between letters which are graphically really similar.

The same has been used in the comparison between Khoji and Voynich alphabets, as shown in Fig. 3 and 4. A binary and cosine distance has been used there, so we compared various alphabets based on the different similarity measurements. Figure 3 shows the similarities and differences between Voynich and Khoji dialect alphabets based on the binary distance. As visible on the figure, the distance is very problematic because all letters must be perfectly centred because of its binary description. In fact, that means that alphabets were converted into series of ones and Zeros where the 1 means black pixel 0 means white pixel and the binary distance has been calculated like the distance between pixels which has don't have the same colour. The alphabets shall be fitted inside its picture. Any shift cause that the binary results would be totally different. That's why we found out this method is not so well. The other similarity measure in our experiment was cosine distance. The darkest colour so for some letters, it's visible that there is a really visual similarity between selected dialects and as mentioned below in the other paragraphs of this paper. All other measurements ([10], [2], [12], [4], [13]) it's possible to combine, so we try also a combination of different measures to create some average weighted measurement of the similarities between the alphabets. There's a lot of other combinations and variable factors, that can be used to set up the cost functions for measurements of the similarities amongst the alphabets. With this kind of experiment, we have a meet a lot of problems, for example, graphical representation of the picture it must be all the same quality, it must be, if possible, be with a certain level of resolution, if possible the alphabets should be written in the same thickness as alphabet from another dialect etc. All those problems influence the recognition of the similarity between the alphabet in a very sensitive way, so if we would like to use this kind of compassion between alphabets. Then be must very carefully consider all those facts and very carefully prepare all data for mutual comparison, which is quite time-consuming and simply complicated.

That's why we have used the second approach, which is focused again on the image similarity, however, this time the comparison is not based on the binary or graphical format of the alphabets point of view, but on the objects attributes in the picture. That means each picture has attributes like shape, brightness contrast and so on and all those attributes describe the object whenever it is lying inside the picture. So this method seems to be much more reliable for comparison of the different alphabets between different dialects in order to find the most similar alphabets or dialects. The principle of the image corresponding points is described in the Fig. 5. There you can clearly see one selected alphabet which is covered by the circles of different colours, with a green dot inside and those cycles basically describes attributes of that of the object. In fact, instead of the binary description of which is a huge and long string of the zeros in one's, in the case of the image corresponding points, we can get very easily only a few numbers which clearly describe the main attributes of a given object. Because all alphabets are prepared in the black colour on the white background, then it's simple for this method to describe alphabets and this method is focused only on the attitudes of that black alphabets. The number of the corresponding points can be adjustable, that means, you can adjust how much those points shall be calculated to take into consideration. In our experiments were set maximal number of points for each letter that means each other has been described by unique corresponding points. This method has again been tested only on the Voynich manuscript in order to check whether this method is able to identify the same letters clearly. That means letters points with the highest appearance of the most same or similar points were evaluated highest. Please remind that those numbers don't mean percentage similarity like in the previous pictures but mean the absolute value of the image corresponding points. You can also see from the Fig. 5 that some letters are similar to another because they visibly contain parts of as a subset of another letter. For example, you can see very easily on the last position on the x-axis the last letter has similarity 196 with another letter on the y-axis. The letter on the y-axis is very similar. You can find more examples thereof that the similarity measurement is much more reliable than the previously tested. The Fig. 5 can also be visualized as a network on Fig. 7a and 7b. You can see a graph with vertices-alphabets from the Voynich manuscript and links that are the relations between those letters. The links between vertices represent the similarity like the alphabets at Fig. 7a. It shows the strength of the similarity between alphabets by colours. These pictures don't correspond exactly with Fig. 6, it has been generated from other experimentations where the colours represent the similarity upon the user-set threshold. This picture can show clearly not the only similarity between the letters in the alphabet but also its strength. This is also visible in Fig. 7b. Thes methods seem to be promising and helpful in the comparison between different dialects. It has been tested only on the Voynich alphabets to verify if it's work or not. The same we can see in Fig. 8a and 8b). The size and colours of the letters represent the importance of the letters in the alphabet and text. Figure 8b) shows basically the same; the difference is that we use the community visualization based on the similarity between the letters in the alphabets. It's clearly visible that this method also nicely group the letter that is very similar from a graphical point of view and seems to be a mutation of one from the other. Based on such results, in the table and the graph, it seems to be that it is possible to use this method to describe the letters from the alphabet amongst them. In the last experiment have been tested Voynich script and Khoji dialect in order to show how those letters are similar to each other or not. From selected results, which are reported on Fig. 9 and 10 is visible that our methods have been able to identify not only the alphabets which are similar to each other but also some of Khoji has been included into Voynich manuscript because they are similar. This is better visible in Fig. 10. Six examples show here how are method based on the image corresponding points associated with the alphabets level of similarity. It is clearly visible that those letters are visually quite similar to its complexity point of view.

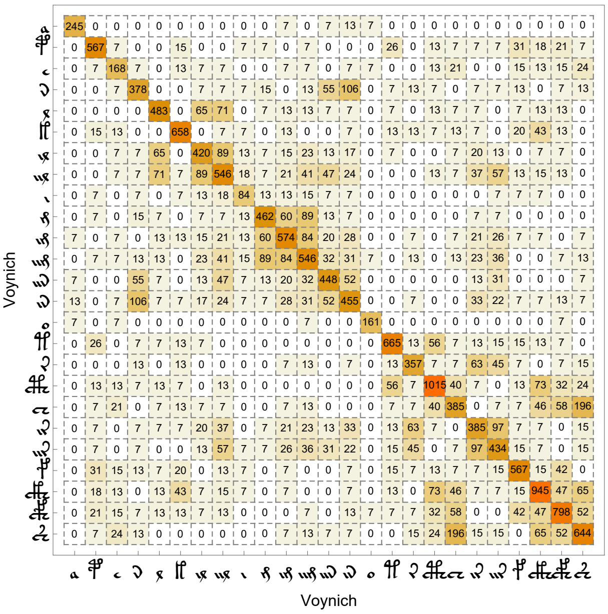
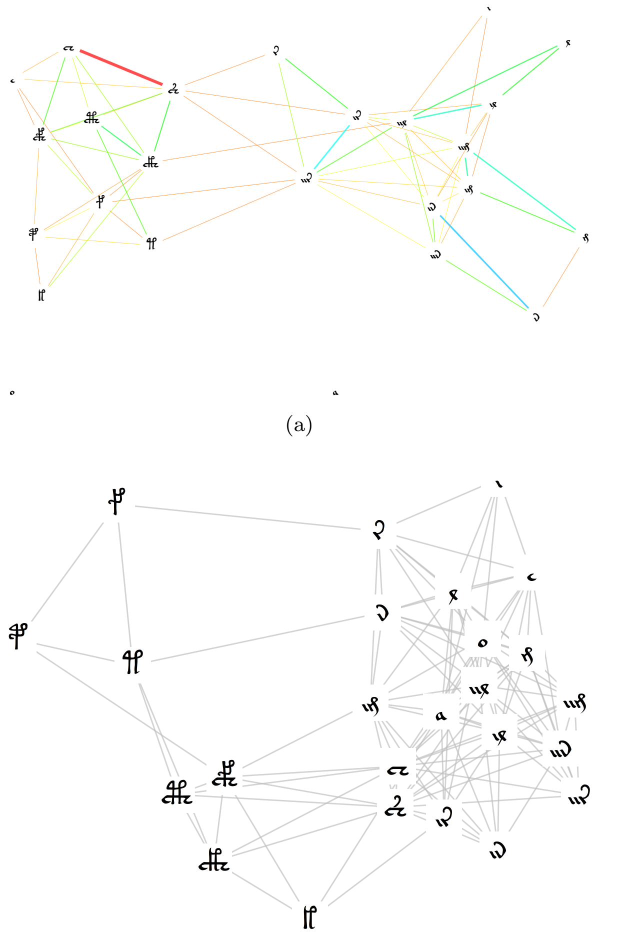


Figure 6: Similarity between alphabets from Voynich based on image corresponding points description. Note that numbers does not represents similarity in %, but an absolute value of the same image corresponding points description between alphabets.

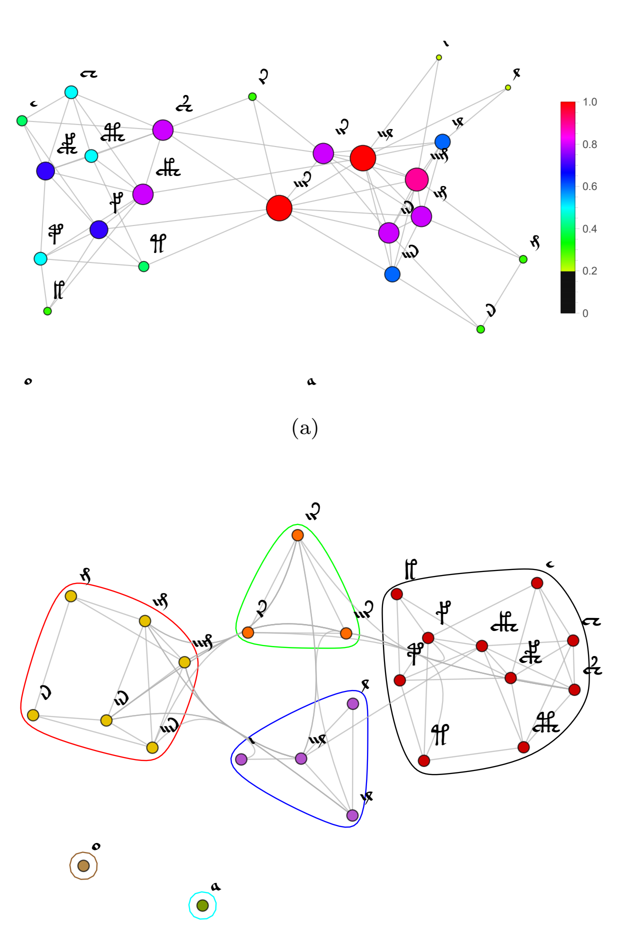
5 Conclusion

In this paper we discuss the possibility on the Voynich manuscripts analyses with using of the different methods from the image similarity on the beginning with primitive methods based on the binary distance measure as well as the others *EuclideanDistance, SquaredEuclideanDistance, NormalizedSquaredEuclideanDistance, ManhattanDistance, CosineDistance, Corre-*


(a)

(b)

Figure 7: An alternative visualization of the similarity between alphabets from Voynich based on image corresponding points description.The similarity and hierarchical relations amongst the alphabets is obvious. The color as in a) can represent strength of the relation.



(a)

(b)

Figure 8: An alternative visualization of the similarity between alphabets from Voynich based on image corresponding points description. The similarity and hierarchical relations amongst the alphabets is obvious. The color as in a) can represent strength of the relation.

lationDistance, RootMeanSquare up to the most advanced based on image corresponding points (that is object position and orientation free in the picture).

In our experiments, we have used a few approaches from the simplest to the most complex. In the first experiment, we have tried to describe the letters of the manuscript just pictures in the binary regime and compare mutually by difference measurement of the distance between those pictures like binary distance, cosine distance and many others. The results we have got shows, that this method is working, however not very well because all letters from alphabets must be generated in the same format, must be centred in the picture on the same positions, must have the same colours, the same picture resolution and so on. So this method is very sensitive for the settings because other dialects and their alphabets can be written in a different form with different thickness and different positions inside the picture, and all this must be corrected. It is possible to use this method with very low efficiency and performance. This method has been tested on the Voynich as well as on the Khoji dialect.

Another method was tested both on the Voynich handwriting alphabet and on the alphabet of the Khoji dialect. This method, we tried, was a description of individual letters of the Voynich alphabet and dialect alphabet, based on the so-called image corresponding point which is a description of an object in the image-based, on important elements of this object like significant points of curvature, of brightness, contrast and other important attributes that uniquely describe the object. In this case, we arrive at results that proved to be very interesting and demonstrate the applicability of this method to describe the letters of the alphabet and measure their similarity both between each other and between different alphabets, regardless of its orientation and position in the picture.

We have also tried the visualization of the Voynich alphabet in the form of graph and community graph, which is shown in Fig. 7. Here we see a graph with coloured vertices that belong to individual characters and their colour and thickness shows the importance of the letter of this alphabet in the text and its greatest

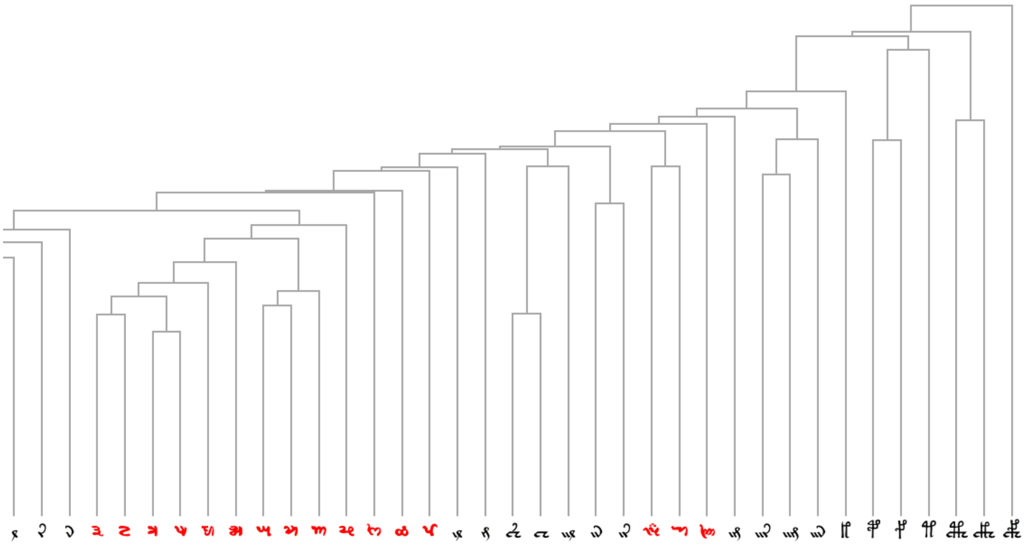


Figure 9: The Khoji dendrogram constructed from the hierarchical clustering of Voynich and Khoji.

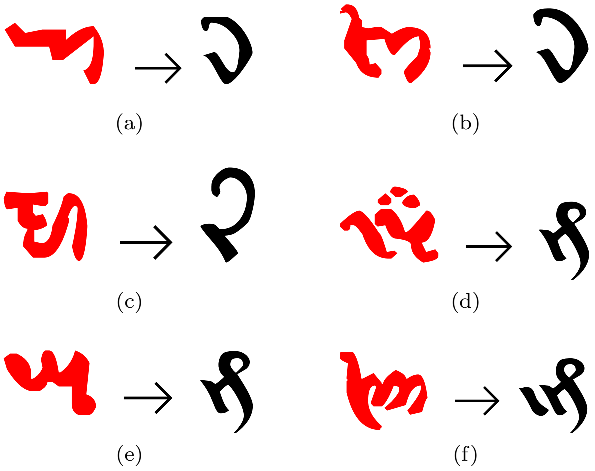


Figure 10: Selected examples of the similar letters between Khoji dialect and Voynich based on image corresponding points.

link similarity with the other letters. In Fig. 7b we see a community graph, where the appropriate algorithm for displaying communities in the graph showed the communities of these letters based on the similarity measurement. From this graph, it is clear that the algorithm recognized letters that are similar in some way letters and created in this case 4 Communities, i.e. 4 clusters of letters that are visually very similar or are the mutations of their say group members.

At the end of our experiments, we have tried another way of visualization using hierarchical clustering in the form of the so-called dendrogram. Figure 9 is an example of such a dendrogram whose selected elements can be seen in detail in the next Fig. 10. We must realize that this similarity is taken in terms of the similarity of the vector which contain just the data of the image corresponding points and therefore these objects may be rotated or may not be 100% similar. These characters of the Khoji alphabet, which have been assigned to the relevant characters of the Voynich alphabet, are very similar in their complexity in that way of writing and overall appearance. Alphabets of different dialects which may not necessarily lead directly to find what dialect Voynich belongs to, but also to the fact that it will be able to measure with it to show the gradual development of the dialect from its first version to the modern version as well. All this in the context of the Voynich manuscript can be a very important part of its research.

Acknowledgement: The following grants are acknowledged for the financial support provided for this research: Grant of SGS No. SP2020/78, VSB Technical University of Ostrava.

References

[1] Amancio, D. R., Altmann, E. G., Rybski, D., Oliveira Jr, O. N., and Costa, L. d. F. Probing the statistical properties of unknown texts: application to the voynich manuscript. *PLoS One 8*, 7 (2013), e67310.

[2] Andres, J., Benešová, M., Kubáček, L., and Vrbková, J. Methodological note on the fractal analysis of texts. *Journal of Quantitative Linguistics 19*, 1 (2012), 1–31.

[3] Brumbaugh, R. S. Botany and the voynich" roger bacon" manuscript once more. *Speculum 49*, 3 (1974), 546–548.

[4] Chaabouni, A., Boubaker, H., Kherallah, M., Alimi, A. M., and El Abed, H. Fractal and multi-fractal for arabic offline writer identification. In *2010 20th International Conference on Pattern Recognition* (2010), IEEE, pp. 3793–3796.

[5] Currier, P. Papers on the voynich manuscript. In *New Research on the Voynich Manuscript: Proceedings of a Seminar, 30 November 1976* (2011).

[6] d'Imperio, M. E. The voynich manuscript: An elegant enigma. Tech. rep., NATIONAL SECURITY AGENCY/CENTRAL SECURITY SERVICE FORT GEORGE G MEADE MD, 1978.

[7] Kennedy, G., and Churchill, R. *The Voynich manuscrip.* Orion Publishing Company, 2004.

[8] Kennedy, G., and Churchill, R. *The Voynich manuscript: The mysterious code that has defied interpretation for centuries.* Simon and Schuster, 2006.

[9] Landini, G. Evidence of linguistic structure in the voynich manuscript using spectral analysis. *Cryptologia 25*, 4 (2001), 275–295.

[10] Martin, J. R. Text and clause: Fractal resonance. *Text-Interdisciplinary Journal for the Study of Discourse 15*, 1 (1995), 5–42.

[11] Montemurro, M. A., and Zanette, D. H. Keywords and co-occurrence patterns in the voynich manuscript: An information-theoretic analysis. *PloS one 8*, 6 (2013), e66344.

[12] Tang, Y. Y., Ma, H., Mao, X., Liu, D., and Suen, C. Y. A new approach to document analysis based on modified fractal signature. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (1995), vol. 2, IEEE, pp. 567–570.

[13] Tucker, A. O., and Janick, J. Identification of phytomorphs in the voynich codex. *Horticultural Reviews 44* (2016), 1–64.

[14] Tucker, A. O., and Talbert, R. *A preliminary analysis of the botany, zoology, and mineralogy of the Voynich manuscript.* Herb News, 2013.

[15] Zelinka, I., Zmeskal, O., Windsor, L., and Cai, Z. Unconventional methods in voynich manuscript analysis. *Mendel 25* (2019), 1–14.