

Clustering by Similarity

1. Calculate a max-min similarity-relation $R = [a_{ij}]$
2. Set $a_{ij} = 0$ for all $a_{ij} < \alpha$ and $i = j$
3. Select s and t so that $a_{ij} = \max\{a_{ij} \mid i < j \text{ and } i, j \in I\}$. When tie, select one of these pairs at random
 WHILE $a_{st} \neq 0$ DO put s and t into the same cluster $C = \{s, t\}$ ELSE [4.] ELSE all indices $\in I$ into separated clusters and STOP
4. Choose $u \in I \setminus C$ so that $\sum_{i \in C} a_{iu} = \max_{j \in I \setminus C} \{ \sum_{i \in C} a_{ij} \mid a_{ij} \neq 0 \}$ When a tie, select one such u at random. WHILE such a u exists, put u into $C = \{s, t, u\}$ and REPEAT [4.]
5. Let $I = I \setminus C$ and GOTO [3.]

I chose the following russian letters for test:

Р, Ю, И, Ы, М, Н, К, П, Ф, О

Starting with

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.6
Ю	0.9	1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.5
И	0.1	0.1	1	0.7	0.5	0.8	0.7	0.4	0.1	0.2
Ы	0.1	0.1	0.7	1	0.6	0.5	0.3	0.2	0.1	0.1
М	0.1	0.1	0.5	0.6	1	0.3	0.3	0.3	0.1	0.2
Н	0.1	0.1	0.8	0.5	0.3	1	0.6	0.5	0.1	0.1
К	0.1	0.1	0.7	0.3	0.3	0.6	1	0.7	0.1	0.1
П	0.1	0.1	0.4	0.2	0.3	0.5	0.7	1	0.1	0.1
Ф	0.3	0.3	0.1	0.1	0.1	0.1	0.1	0.1	1	0.4
О	0.6	0.5	0.2	0.1	0.2	0.1	0.1	0.1	0.4	1

So we'll modify the matrix:

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Ю	0.9	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
И	0.1	0.1	1	0.8	0.1	0.1	0.2	0.2	0.2	0.8
Ы	0.1	0.1	0.8	1	0.1	0.1	0.2	0.1	0.1	0.8
М	0.1	0.1	0.1	0.1	1	0.9	0.1	0.1	0.1	0.1
Н	0.1	0.1	0.1	0.1	0.9	1	0.1	0.1	0.1	0.1
К	0.1	0.1	0.2	0.2	0.1	0.1	1	0.2	0.2	0.2
П	0.1	0.1	0.2	0.1	0.1	0.1	0.2	1	0.6	0.1
Ф	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.6	1	0.1
О	0.1	0.1	0.8	0.8	0.1	0.1	0.2	0.1	0.1	1

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Ю	0.9	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
И	0.1	0.1	1	0.8	0.1	0.1	0.2	0.2	0.2	0.8
Ы	0.1	0.1	0.8	1	0.1	0.1	0.2	0.2	0.2	0.8
М	0.1	0.1	0.1	0.1	1	0.9	0.1	0.1	0.1	0.1
Н	0.1	0.1	0.1	0.1	0.9	1	0.1	0.1	0.1	0.1
К	0.1	0.1	0.2	0.2	0.1	0.1	1	0.2	0.2	0.2
П	0.1	0.1	0.2	0.2	0.1	0.1	0.2	1	0.6	0.2
Ф	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.6	1	0.2
О	0.1	0.1	0.8	0.8	0.1	0.1	0.2	0.2	0.2	1

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	1	0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Ю	0.9	1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
И	0.1	0.1	1	0.8	0.1	0.1	0.2	0.2	0.2	0.8
Ы	0.1	0.1	0.8	1	0.1	0.1	0.2	0.2	0.2	0.8
М	0.1	0.1	0.1	0.1	1	0.9	0.1	0.1	0.1	0.1
Н	0.1	0.1	0.1	0.1	0.9	1	0.1	0.1	0.1	0.1
К	0.1	0.1	0.2	0.2	0.1	0.1	1	0.2	0.2	0.2
П	0.1	0.1	0.2	0.2	0.1	0.1	0.2	1	0.6	0.2
Ф	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.6	1	0.2
О	0.1	0.1	0.8	0.8	0.1	0.1	0.2	0.2	0.2	1

1) So then lets change all values that are < 0.55 and all values on main diagonal to 0 .

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	0	0.9	0	0	0	0	0	0	0	0
Ю	0.9	0	0	0	0	0	0	0	0	0
И	0	0	0	0.8	0	0	0	0	0	0.8
Ы	0	0	0.8	0	0	0	0	0	0	0.8
М	0	0	0	0	0	0.9	0	0	0	0
Н	0	0	0	0	0.9	0	0	0	0	0
К	0	0	0	0	0	0	0	0	0	0
П	0	0	0	0	0	0	0	0	0.6	0
Ф	0	0	0	0	0	0	0	0.6	0	0
О	0	0	0.8	0.8	0	0	0	0	0	0

2-4) find maximum value with coordinates (5,6) =0.9 and add it to cluster 1 so there are М, Н in cluster 1.

	Р	Ю	И	Ы	М	Н	К	П	Ф	О
Р	0	0.9	0	0	0	0	0	0	0	0
Ю	0.9	0	0	0	0	0	0	0	0	0
И	0	0	0	0.8	0	0	0	0	0	0.8
Ы	0	0	0.8	0	0	0	0	0	0	0.8
М	0	0	0	0	0	0.9	0	0	0	0
Н	0	0	0	0	0.9	0	0	0	0	0
К	0	0	0	0	0	0	0	0	0	0
П	0	0	0	0	0	0	0	0	0.6	0
Ф	0	0	0	0	0	0	0	0.6	0	0
О	0	0	0.8	0.8	0	0	0	0	0	0

5-6) Then we exclude those letters from table and continue the process till all the letters are in clusters.

	Р	Ю	И	Ы	К	П	Ф	О
Р	0	0.9	0	0	0	0	0	0
Ю	0.9	0	0	0	0	0	0	0
И	0	0	0	0.8	0	0	0	0.8
Ы	0	0	0.8	0	0	0	0	0.8
К	0	0	0	0	0	0	0	0
П	0	0	0	0	0	0	0.6	0
Ф	0	0	0	0	0	0.6	0	0
О	0	0	0.8	0.8	0	0	0	0

And so on...

So we have the following clusters: {М, Н}, {Р, Ю}, {И, Ы, О}, {П, Ф}, {К}.