

# Text similarity in Finnic oral folk poetry

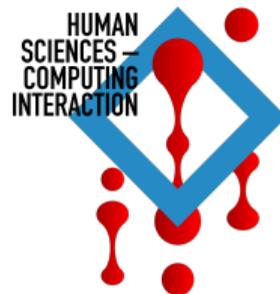
## Towards a large-scale quantitative analysis

Maciej Janicki

October 20, 2022



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

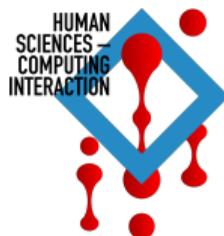


# THE FILTER PROJECT

Formulaic intertextuality, thematic networks and poetic variation  
across regional cultures of Finnic oral poetry

A cooperation of:

- 1 **Finnish Literature Society**  
Kati Kallio, Jukka Saarinen
- 2 **University of Helsinki, HSCI group**  
Eetu Mäkelä, Maciej Janicki
- 3 **Estonian Literary Museum**  
Mari Sarv, Liina Saarlo



| | | | | | | | | | Estonian Literary Museum

# THE CORPUS

## Text collections:

- 1 Suomen Kansan Vanhat Runot (SKVR)
- 2 Eesti Regilaulude Andmebaas (ERAB)
- 3 SKS's corpus of unpublished poems (JR)
- 4 Literary works: e.g. Kalevala, Kanteletar, ...

## Characteristics:

- Finnic tetrameter
- mostly 19th-early 20th century
- size: **275,274** poems, 4,388,348 verses.

Lappalaiñe on kyttösilmä  
Piti viikoista vihuva,  
Kauvon aijasta katsetta.  
Huol'itti tulista nuolda  
Tul'izilla jändevillä,  
Peäškön pieñillä sulilla,  
Varposen vivuttsimilla.  
SKVR I1 17

Mõistke, mõistke, mehed noored  
Teadke, teadke naesed targad  
Arvake küla õeksed  
Mis see seal mäella kasvab  
Tamme seal mäella kasvab  
EÜS VII 313 (67)

# GOAL: ALIGNMENT

Yht' ei kuttusun Lemmingäistä.  
\*Rujot (ne) reillä reissuaabi,  
Rimmat rattšahin ajeli,  
[20\\_Sogiat venozin soudi.\\*](#)

Lemmingäin on poiga\_piilo  
Pillojah on piilemässä,#10  
Pahojah pagenemassa.#11  
"Hoib om moamo, kandajañi,  
[25\\_Armas maijon andajañi,](#)  
Ihalan imettäjäni,  
Ettsib om miul pelvoi paid[a],  
Ennembä neidona kuvottu,  
Kassabapeän#12 on kalkuteltu,  
[30\\_Kannabas paloni\\_paid\[a\]."](#)

<http://runoregi.rahtiapp.fi/poemdiff?nro1=skvr07108030&nro2=skvr07108080>

15 Kuttšu veri-sogeat,  
Ruiot re'ellä rembutteli,  
Rimmat rattšahin ajeli,  
Sogeat venosin souti,  
Yht' ei kuttšu Lemmingästä.  
[20\\_Lemmingäne on piilopoiga](#)  
Pilloja on piilemässä,  
Pahoja pagenemassa.  
"Hoi on moammoni, kantajani,  
Armas maion antajani,  
[25\\_Ihala imettäjäni,](#)

Tuos miull' sot'isobani,  
Kannas paloini-paita!"

- 1 align similar texts line-by-line
- 2 compute similarity measures summarizing the alignment
- 3 automatically find pairs of similar texts in the corpus
- 4 get large-scale overviews of the identified similarities

# SIMILARITY MEASURES

|                           |                             |     |
|---------------------------|-----------------------------|-----|
| kuttsu_verisogeat         | yht_ei_kuttsun_lemmingäistä | 0   |
| ruiot_reellä_rebutteli    | rujot_ne_reillä_reissuaabi  | 0   |
| rammat_rattsahin_ajeli    | rammat_rattsahin_ajeli      | 1   |
| sogeat_venosin_souti      | sogiat_venozin_soudi        | .71 |
| yht_ei_kuttsu_lemmingästä |                             | 0   |
| lemmingäne_on_piilopoiga  | lemmingäin_on_poiga_piilo   | .84 |
| piiloja_on_piilemässä     | pillojah_on_piilemässä      | .86 |
| pahoja_pagenemassa        | pahojah_pagenemassa         | .93 |

$$s_{raw} = 4.34$$
$$s_l = \frac{4.34}{8} \approx 0.54$$
$$s_r = \frac{4.34}{7} = 0.62$$
$$s = \frac{2 \cdot 4.34}{8 + 7} \approx 0.58$$

yht\_ei\_kuttsun\_lemmingäistä  
 rujot\_ne\_reillä\_reissuaabi  
 rammat\_rattsahin\_ajeli  
 sogiat\_venozin\_souti  
 lemmingäin\_on\_poiga\_piilo  
 pillojah\_on\_piilemässä  
 pahojah\_pagenemassa

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

kuttsu\_verisogeat  
 ruiot\_reellä\_rembutteli  
 rammat\_rattsahin\_ajeli  
 sogeat\_venosin\_souti  
 yht\_ei\_kuttsu\_lemmingästä  
 lemmingäne\_on\_piilopoiga  
 piiloja\_on\_piilemässä  
 pahoja\_pagenemassa

.88

1

.71

.84

.52

.57

.86

.93

|                           |   | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                           | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0 |                             |                            |                        |                      |                           |                        |                     |
| ruiot_reellä_rembutteli   | 0 |                             |                            |                        |                      |                           |                        |                     |
| rammat_rattsahin_ajeli    | 0 |                             |                            | 1                      |                      |                           |                        |                     |
| sogeat_venosin_souti      | 0 |                             |                            |                        | .71                  |                           |                        |                     |
| yht_ei_kuttsu_lemmingästä | 0 | .88                         |                            |                        |                      |                           |                        |                     |
| lemmingäne_on_piilopoiga  | 0 |                             |                            |                        |                      | .84                       | .52                    |                     |
| piiloja_on_piilemässä     | 0 |                             |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0 |                             |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           |   | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | piilojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                           | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0 |                             |                            |                        |                      |                           |                        |                     |
| rammat_rattsahin_ajeli    | 0 |                             |                            | 1                      |                      |                           |                        |                     |
| sogeat_venosin_souti      | 0 |                             |                            |                        | .71                  |                           |                        |                     |
| yht_ei_kuttsu_lemmingästä | 0 | .88                         |                            |                        |                      |                           |                        |                     |
| lemmingäne_on_piilopoiga  | 0 |                             |                            |                        |                      | .84                       | .52                    |                     |
| piiloja_on_piilemässä     | 0 |                             |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0 |                             |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           |   | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                           | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0 |                             |                            | 1                      |                      |                           |                        |                     |
| sogeat_venosin_souti      | 0 |                             |                            |                        | .71                  |                           |                        |                     |
| yht_ei_kuttsu_lemmingästä | 0 | .88                         |                            |                        |                      |                           |                        |                     |
| lemmingäne_on_piilopoiga  | 0 |                             |                            |                        |                      | .84                       | .52                    |                     |
| piiloja_on_piilemässä     | 0 |                             |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0 |                             |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           |   | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | piilojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                           | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0 | 0                           | 0                          | 1                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0 |                             |                            | .71                    |                      |                           |                        |                     |
| yht_ei_kuttsu_lemmingästä | 0 | .88                         |                            |                        |                      |                           |                        |                     |
| lemmingäne_on_piilopoiga  | 0 |                             |                            |                        | .84                  | .52                       |                        |                     |
| piiloja_on_piilemässä     | 0 |                             |                            |                        | .57                  | .86                       |                        |                     |
| pahoja_pagenemassa        | 0 |                             |                            |                        |                      |                           | .93                    |                     |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| 0                         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0                           | 0                          | 0                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0                           | 0                          | 0                      | 1                    | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingästä | 0                           | .88                        |                        |                      |                           |                        |                     |
| lemmingäne_on_piilopoiga  | 0                           |                            |                        |                      | .84                       | .52                    |                     |
| piiloja_on_piilemässä     | 0                           |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0                           |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| 0                         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0                           | 0                          | 0                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0                           | 0                          | 0                      | 1                    | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingästä | 0                           | .88                        | .88                    | 1                    | 1.71                      | 1.71                   | 1.71                |
| lemmingäne_on_piilopoiga  | 0                           |                            |                        |                      | .84                       | .52                    |                     |
| piiloja_on_piilemässä     | 0                           |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0                           |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| 0                         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0                           | 0                          | 0                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0                           | 0                          | 0                      | 1                    | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingästä | 0                           | .88                        | .88                    | 1                    | 1.71                      | 1.71                   | 1.71                |
| lemmingäne_on_piilopoiga  | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 2.55                |
| piiloja_on_piilemässä     | 0                           |                            |                        |                      | .57                       | .86                    |                     |
| pahoja_pagenemassa        | 0                           |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_soudi | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| 0                         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0                           | 0                          | 0                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0                           | 0                          | 0                      | 1                    | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingästä | 0                           | .88                        | .88                    | 1                    | 1.71                      | 1.71                   | 1.71                |
| lemmingäne_on_piilopoiga  | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 2.55                |
| piiloja_on_piilemässä     | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 3.41                |
| pahoja_pagenemassa        | 0                           |                            |                        |                      |                           |                        | .93                 |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                            |   | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|----------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                            | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat          | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli    | 0 | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli     | 0 | 0                           | 0                          | 1                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti       | 0 | 0                           | 0                          | 1                      | 1.71                 | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingäistä | 0 | .88                         | .88                        | 1                      | 1.71                 | 1.71                      | 1.71                   | 1.71                |
| lemmingäne_on_piilopoiga   | 0 | .88                         | .88                        | 1                      | 1.71                 | 2.55                      | 2.55                   | 2.55                |
| piiloja_on_piilemässä      | 0 | .88                         | .88                        | 1                      | 1.71                 | 2.55                      | 3.41                   | 3.41                |
| pahoja_pagenemassa         | 0 | .88                         | .88                        | 1                      | 1.71                 | 2.55                      | 3.41                   | 4.34                |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

|                           | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | piilojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
|                           | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| kuttsu_verisogeat         | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| ruiot_reellä_rembutteli   | 0                           | 0                          | 0                      | 0                    | 0                         | 0                      | 0                   |
| rammat_rattsahin_ajeli    | 0                           | 0                          | 0                      | 1                    | 1                         | 1                      | 1                   |
| sogeat_venosin_souti      | 0                           | 0                          | 0                      | 1                    | 1.71                      | 1.71                   | 1.71                |
| yht_ei_kuttsu_lemmingästä | 0                           | .88                        | .88                    | 1                    | 1.71                      | 1.71                   | 1.71                |
| lemmingäne_on_piilopoiga  | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 2.55                |
| piiloja_on_piilemässä     | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 3.41                |
| pahoja_pagenemassa        | 0                           | .88                        | .88                    | 1                    | 1.71                      | 2.55                   | 3.41                |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

$$s_{\text{raw}} = 4.34$$

$$s_l = \frac{4.34}{8} \approx 0.54$$

$$s_r = \frac{4.34}{7} = 0.62$$

$$s = \frac{2 \cdot 4.34}{8 + 7} \approx 0.58$$

|                           | 0 | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | piilojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| kuttsu_verisogeat         | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| ruiot_reellä_rembutteli   | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| rammat_rattsahin_ajeli    | ↑ | ↑                           | ↑                          | ↘                      | ←                    | ←                         | ←                      | ←                   |
| sogeat_venosin_souti      | ↑ | ↑                           | ↑                          | ↑                      | ↘                    | ←                         | ←                      | ←                   |
| yht_ei_kuttsu_lemmingästä | ↑ | ↘                           | ←                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| lemmingäne_on_piilopoiga  | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↘                         | ←                      | ←                   |
| piiloja_on_piilemässä     | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↘                      | ←                   |
| pahoja_pagenemassa        | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↘                   |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

↑  
←  
↘

$$s_{\text{raw}} = 4.34$$

$$s_l = \frac{4.34}{8} \approx 0.54$$

$$s_r = \frac{4.34}{7} = 0.62$$

$$s = \frac{2 \cdot 4.34}{8 + 7} \approx 0.58$$

|                           | 0 | yht_ei_kuttsun_lemmingäistä | rujot_ne_reillä_reissuaabi | rammat_rattsahin_ajeli | sogiat_venozin_souti | lemmingäin_on_poiga_piilo | pillojah_on_piilemässä | pahojah_pagenemassa |
|---------------------------|---|-----------------------------|----------------------------|------------------------|----------------------|---------------------------|------------------------|---------------------|
| kuttsu_verisogeat         | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| ruiot_reellä_rembutteli   | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| rammat_rattsahin_ajeli    | ↑ | ↑                           | ↑                          | ↘                      | ←                    | ←                         | ←                      | ←                   |
| sogeat_venosin_souti      | ↑ | ↑                           | ↑                          | ↑                      | ↘                    | ←                         | ←                      | ←                   |
| yht_ei_kuttsu_lemmingästä | ↑ | ↙                           | ←                          | ↑                      | ↑                    | ↑                         | ↑                      | ↑                   |
| lemmingäne_on_piilopoiga  | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↘                         | ←                      | ←                   |
| piiloja_on_piilemässä     | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↘                      | ←                   |
| pahoja_pagenemassa        | ↑ | ↑                           | ↑                          | ↑                      | ↑                    | ↑                         | ↑                      | ↘                   |

$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\}$$

↑  
←  
↖

$$s_{\text{raw}} = 4.34$$

$$s_l = \frac{4.34}{8} \approx 0.54$$

$$s_r = \frac{4.34}{7} = 0.62$$

$$s = \frac{2 \cdot 4.34}{8 + 7} \approx 0.58$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$D = \begin{array}{c} \text{p} \\ \text{o} \\ \text{t} \\ \text{t} \\ \text{e} \\ \text{r} \\ \text{y} \end{array} \begin{array}{cccccc} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ \left[ \begin{array}{cccccc} 0 & 1 & & & & & \\ 0 & & 1 & & & & \\ 0 & & & & 1 & & \\ 0 & & & & 1 & & \\ 0 & & & 1 & & & \\ 0 & & & & & 1 & \\ 0 & & & & & & 1 \end{array} \right] \end{array}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 1$$
$$D_{i,2} = \begin{matrix} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ [1 & 0 & 0 & 0 & 0 & 0 & 0] \end{matrix}$$
$$d_{i,j} = \max \left\{ \begin{matrix} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{matrix} \right\} = \max \left\{ \begin{matrix} 0 \\ d_{i,j-1} \\ 0 + d_{i,j} \end{matrix} \right\}$$
$$D_{i,2} = \text{cummax}(D_{i,2,:})$$
$$= [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

$$D = \begin{matrix} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ \text{p} & [0 & 1 & & & & ] \\ \text{o} & [0 & & 1 & & & ] \\ \text{t} & [0 & & & 1 & & ] \\ \text{t} & [0 & & & & 1 & ] \\ \text{e} & [0 & & 1 & & & ] \\ \text{r} & [0 & & & & & 1] \\ \text{y} & [0 & & & & & & 1] \end{matrix}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 1$$
$$D_{i,2:} = \begin{array}{c} \text{p} \quad \text{o} \quad \text{e} \quad \text{t} \quad \text{r} \quad \text{y} \\ [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0] \end{array}$$
$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\} = \max \left\{ \begin{array}{l} 0 \\ d_{i,j-1} \\ 0 + d_{i,j} \end{array} \right\}$$
$$D_{i,2:} = \text{cummax}(D_{i,2:})$$
$$= [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

$$D = \begin{array}{c} \text{p} \quad \text{o} \quad \text{e} \quad \text{t} \quad \text{r} \quad \text{y} \\ \text{p} \left[ \begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right. \\ \text{o} \left[ \begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right. \\ \text{t} \left[ \begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right. \\ \text{t} \left[ \begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \right. \\ \text{e} \left[ \begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{array} \right. \\ \text{r} \left[ \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{array} \right. \\ \text{y} \left[ \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{array} \right. \end{array} \right]$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 1$$
$$D_{i,2} = \begin{array}{cccccc} & p & o & e & t & r & y \\ [1 & 0 & 0 & 0 & 0 & 0 & 0] \end{array}$$
$$d_{i,j} = \max \left\{ \begin{array}{l} d_{i-1,j} \\ d_{i,j-1} \\ d_{i-1,j-1} + d_{i,j} \end{array} \right\} = \max \left\{ \begin{array}{l} 0 \\ d_{i,j-1} \\ 0 + d_{i,j} \end{array} \right\}$$
$$D_{i,2} = \text{cummax}(D_{i,2:})$$
$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$D = \begin{array}{cccccc} & p & o & e & t & r & y \\ p & 0 & 1 & 1 & 1 & 1 & 1 \\ o & 0 & & 1 & & & \\ t & 0 & & & 1 & & \\ t & 0 & & & & 1 & \\ e & 0 & & 1 & & & \\ r & 0 & & & & 1 & \\ y & 0 & & & & & 1 \end{array}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 2$$

$$D_{i,2} = \begin{array}{c} \text{p o e t r y} \\ [0 \ 1 \ 0 \ 0 \ 0 \ 0] \end{array}$$

$$D_{i,2}^* = \text{fmax} \left\{ \begin{array}{c} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 1 \ 1 \ 1 \ 1 \ 1] \end{array} + D_{i,2} \right\}$$

$$= \text{fmax} \left\{ \begin{array}{c} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 2 \ 1 \ 1 \ 1 \ 1] \end{array} \right\}$$

$$= [1 \ 2 \ 1 \ 1 \ 1 \ 1]$$

$$D_{i,2} = \text{cummax}(D_{i,2}^*) = [1 \ 2 \ 2 \ 2 \ 2 \ 2]$$

$$D = \begin{array}{c} \text{p o e t r y} \\ \text{p} \\ \text{o} \\ \text{t} \\ \text{t} \\ \text{e} \\ \text{r} \\ \text{y} \end{array} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & & & & 1 & & \\ 0 & & & & 1 & & \\ 0 & & & 1 & & & \\ 0 & & & & & 1 & \\ 0 & & & & & & 1 \end{bmatrix}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 2$$

$$D_{i,2} = \begin{array}{c} \text{p o e t r y} \\ [0 \ 1 \ 0 \ 0 \ 0 \ 0] \end{array}$$

$$D_{i,2}^* = \text{fmax} \left\{ \begin{array}{c} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 1 \ 1 \ 1 \ 1 \ 1] + D_{i,2} \end{array} \right\}$$

$$= \text{fmax} \left\{ \begin{array}{c} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 2 \ 1 \ 1 \ 1 \ 1] \end{array} \right\}$$

$$= [1 \ 2 \ 1 \ 1 \ 1 \ 1]$$

$$D_{i,2} = \text{cummax}(D_{i,2}^*) = [1 \ 2 \ 2 \ 2 \ 2 \ 2]$$

$$D = \begin{array}{c} \text{p o e t r y} \\ \text{p} \\ \text{o} \\ \text{t} \\ \text{t} \\ \text{e} \\ \text{r} \\ \text{y} \end{array} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & & & 1 & & \\ 0 & & & & 1 & \\ 0 & & 1 & & & \\ 0 & & & & 1 & \\ 0 & & & & & 1 \end{bmatrix}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$\begin{aligned}i &= 2 \\ D_{i,2} &= \begin{array}{cccccc} & p & o & e & t & r & y \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}\end{array} \\ D_{i,2}^* &= \text{fmax} \left\{ \begin{array}{c} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} + D_{i,2} \end{array} \right\} \\ &= \text{fmax} \left\{ \begin{array}{c} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \begin{bmatrix} 0 & 2 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \right\} \\ &= \begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 1 \end{bmatrix} \\ D_{i,2} &= \text{cummax}(D_{i,2}^*) = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}\end{aligned}$$

$$D = \begin{array}{cccccc} & p & o & e & t & r & y \\ \begin{array}{c} p \\ o \\ t \\ t \\ e \\ r \\ y \end{array} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & & & & 1 & & \\ 0 & & & & 1 & & \\ 0 & & & 1 & & & \\ 0 & & & & & 1 & \\ 0 & & & & & & 1 \end{bmatrix}\end{array}$$

# OPTIMIZATION 1: ENTIRE ROWS AT ONCE

$$i = 2$$

$$D_{i,2} = \begin{array}{cccccc} & p & o & e & t & r & y \\ [0 & 1 & 0 & 0 & 0 & 0 & 0] \end{array}$$

$$D_{i,2}^* = \text{fmax} \left\{ \begin{array}{l} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 1 \ 1 \ 1 \ 1 \ 1] + D_{i,2} \end{array} \right\}$$

$$= \text{fmax} \left\{ \begin{array}{l} [1 \ 1 \ 1 \ 1 \ 1 \ 1] \\ [0 \ 2 \ 1 \ 1 \ 1 \ 1] \end{array} \right\}$$

$$= [1 \ 2 \ 1 \ 1 \ 1 \ 1]$$

$$D_{i,2} = \text{cummax}(D_{i,2}^*) = [1 \ 2 \ 2 \ 2 \ 2 \ 2]$$

$$D = \begin{array}{cccccc} & p & o & e & t & r & y \\ p & [0 & 1 & 1 & 1 & 1 & 1] \\ o & [0 & 1 & 2 & 2 & 2 & 2] \\ t & [0 & & & 1 & & ] \\ t & [0 & & & 1 & & ] \\ e & [0 & & 1 & & & ] \\ r & [0 & & & & 1 & ] \\ y & [0 & & & & & 1] \end{array}$$

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$D = \begin{array}{c} \begin{array}{c} p \\ o \\ t \\ t \\ e \\ r \\ y \end{array} \\ \hline \begin{array}{c} f \\ o \\ r \\ g \\ e \\ r \\ y \end{array} \end{array} \begin{array}{c} \begin{array}{c} p \\ o \\ e \\ t \\ r \\ y \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \end{array} \begin{array}{c} \begin{array}{c} o \\ e \\ t \\ r \\ y \end{array} \\ \begin{array}{c} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{array} \end{array} \begin{array}{c} \begin{array}{c} t \\ r \\ y \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{array} \end{array} \begin{array}{c} \begin{array}{c} r \\ y \end{array} \\ \begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{array} \end{array} \begin{array}{c} \begin{array}{c} y \end{array} \\ \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{array} \end{array} \end{array}$$

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{1, 8\}$$

$$D_{I,2} = \begin{array}{c} \text{p o e t r y} \\ \text{p} \\ \text{f} \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned} D_{I,2} &= \text{cummax}(D_{I,2}) \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

$$D = \begin{array}{c} \text{p o e t r y} \\ \text{p} \\ \text{o} \\ \text{t} \\ \text{t} \\ \text{e} \\ \text{r} \\ \text{y} \\ \text{f} \\ \text{o} \\ \text{r} \\ \text{g} \\ \text{e} \\ \text{r} \\ \text{y} \end{array} \begin{array}{c} \text{p o e t r y} \\ \begin{bmatrix} 0 & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \\ & & & & & & 1 \\ & & & & & & & 1 \\ & & & & & & & & 1 \\ \hline & & & & & & & & & 1 \\ & & & & & & & & & & 1 \\ & & & & & & & & & & & 1 \\ & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & 1 \end{bmatrix} \end{array}$$

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{1, 8\}$$

$$D_{I,2} = \begin{array}{c} \text{p} \\ \text{f} \end{array} \begin{array}{c} \text{p} \text{ o} \text{ e} \text{ t} \text{ r} \text{ y} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$\begin{aligned} D_{I,2} &= \text{cummax}(D_{I,2}) \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

$$D = \begin{array}{c} \text{p} \\ \text{o} \\ \text{t} \\ \text{t} \\ \text{e} \\ \text{r} \\ \text{y} \\ \text{f} \\ \text{o} \\ \text{r} \\ \text{g} \\ \text{e} \\ \text{r} \\ \text{y} \end{array} \begin{array}{c} \text{p} \text{ o} \text{ e} \text{ t} \text{ r} \text{ y} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & & 1 & & & \\ 0 & & & 1 & & \\ 0 & & & & 1 & \\ 0 & & 1 & & & \\ 0 & & & & & 1 \\ 0 & & & & & & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & & 1 & & & \\ 0 & & & & 1 & \\ 0 & & & 1 & & \\ 0 & & & & & 1 & 1 \end{bmatrix} \end{array}$$

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{1, 8\}$$

$$D_{I,2} = \begin{array}{c} \text{p} \\ \text{f} \end{array} \begin{array}{c} \text{p} \ \text{o} \ \text{e} \ \text{t} \ \text{r} \ \text{y} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$D_{I,2} = \text{cummax}(D_{I,2}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

|   |   | p | o | e | t | r | y |
|---|---|---|---|---|---|---|---|
| p | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| o | 0 |   | 1 |   |   |   |   |
| t | 0 |   |   |   | 1 |   |   |
| t | 0 |   |   |   | 1 |   |   |
| e | 0 |   |   | 1 |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |
| f | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | 0 |   | 1 |   |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| g | 0 |   |   |   |   |   |   |
| e | 0 |   |   | 1 |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{2, 9\}$$

$$D_{I,2} = \begin{matrix} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ \text{o} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \text{o} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$D_{I,2}^* = \text{fmax} \left\{ \begin{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + D_{I,2} \end{matrix} \right\}$$

$$= \begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D_{I,2} = \text{cummax}(D_{I,2}^*) = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$D = \begin{matrix} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ \text{p} & \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \text{o} & \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ \text{t} & \begin{bmatrix} 0 & & & 1 & & \end{bmatrix} \\ \text{t} & \begin{bmatrix} 0 & & & 1 & & \end{bmatrix} \\ \text{e} & \begin{bmatrix} 0 & & & & 1 & \end{bmatrix} \\ \text{r} & \begin{bmatrix} 0 & & & & & 1 \end{bmatrix} \\ \text{y} & \begin{bmatrix} 0 & & & & & & 1 \end{bmatrix} \\ \text{f} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \text{o} & \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ \text{r} & \begin{bmatrix} 0 & & & & 1 & \end{bmatrix} \\ \text{g} & \begin{bmatrix} 0 & & & & & \end{bmatrix} \\ \text{e} & \begin{bmatrix} 0 & & & 1 & & \end{bmatrix} \\ \text{r} & \begin{bmatrix} 0 & & & & & 1 \end{bmatrix} \\ \text{y} & \begin{bmatrix} 0 & & & & & & 1 \end{bmatrix} \end{matrix}$$

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{2, 9\}$$

$$D_{I,2} = \begin{matrix} & \text{p} & \text{o} & \text{e} & \text{t} & \text{r} & \text{y} \\ \text{o} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \text{o} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$D_{I,2}^* = \text{fmax} \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + D_{I,2} \right\}$$

$$= \begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D_{I,2} = \text{cummax}(D_{I,2}^*) = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

|   |   | p | o | e | t | r | y |
|---|---|---|---|---|---|---|---|
| p | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| o | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| t | 0 |   |   | 1 |   |   |   |
| e | 0 |   |   |   | 1 |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |
| f | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| r | 0 |   |   |   |   | 1 |   |
| g | 0 |   |   |   |   |   |   |
| e | 0 |   |   | 1 |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{2, 9\}$$

$$D_{I,2} = \begin{array}{c} \text{p} \quad \text{o} \quad \text{e} \quad \text{t} \quad \text{r} \quad \text{y} \\ \text{o} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$D_{I,2}^* = \text{fmax} \left\{ \begin{array}{c} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} + D_{I,2} \right\}$$

$$= \begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D_{I,2} = \text{cummax}(D_{I,2}^*) = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

|   |   | p | o | e | t | r | y |
|---|---|---|---|---|---|---|---|
| p | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| o | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| t | 0 |   |   | 1 |   |   |   |
| t | 0 |   |   |   | 1 |   |   |
| e | 0 |   |   | 1 |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |
| f | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| o | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| r | 0 |   |   |   |   | 1 |   |
| g | 0 |   |   |   |   |   |   |
| e | 0 |   |   | 1 |   |   |   |
| r | 0 |   |   |   |   | 1 |   |
| y | 0 |   |   |   |   |   | 1 |

# OPTIMIZATION 2: MULTIPLE ROWS AT ONCE

$$I = \{2, 9\}$$

$$D_{I,2} = \begin{matrix} & \begin{matrix} p & o & e & t & r & y \end{matrix} \\ \begin{matrix} o \\ o \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

$$D_{I,2}^* = \text{fmax} \left\{ \begin{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + D_{I,2} \end{matrix} \right\}$$

$$= \begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D_{I,2} = \text{cummax}(D_{I,2}^*) = \begin{bmatrix} 1 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$D = \begin{matrix} & \begin{matrix} p & o & e & t & r & y \end{matrix} \\ \begin{matrix} p \\ o \\ t \\ t \\ e \\ r \\ y \\ f \\ o \\ r \\ g \\ e \\ r \\ y \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 & 2 \\ 0 & & & 1 & & \\ 0 & & & 1 & & \\ 0 & & 1 & & & \\ 0 & & & & 1 & \\ 0 & & & & & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & & & & 1 & \\ 0 & & & 1 & & \\ 0 & & & & & 1 \\ 0 & & & & & 1 \end{bmatrix} \end{matrix}$$

# BENCHMARKS

Computation times (in seconds) for subsets of SKVR:

| #poems | optimization |        |        |        |
|--------|--------------|--------|--------|--------|
|        | 0            | 1      | 2-CPU  | 2-GPU  |
| 100    | 85.8         | 6.7    | 5.8    | 13.4   |
| 1,000  | 3,261        | 283    | 69.6   | 113    |
| 10,000 | –            | 21,963 | 2,319  | 1,800  |
| 88,078 | –            | –      | 92,850 | 32,036 |

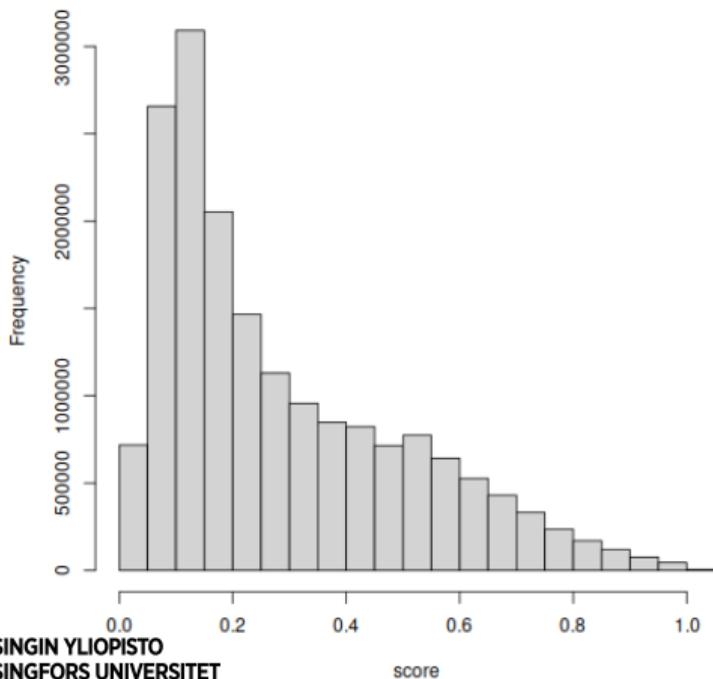
Janicki, M. Optimizing the weighted sequence alignment algorithm for large-scale text similarity computation. In: *Proceedings of NLP4DH 2022*.

# QUANTITATIVE ANALYSIS

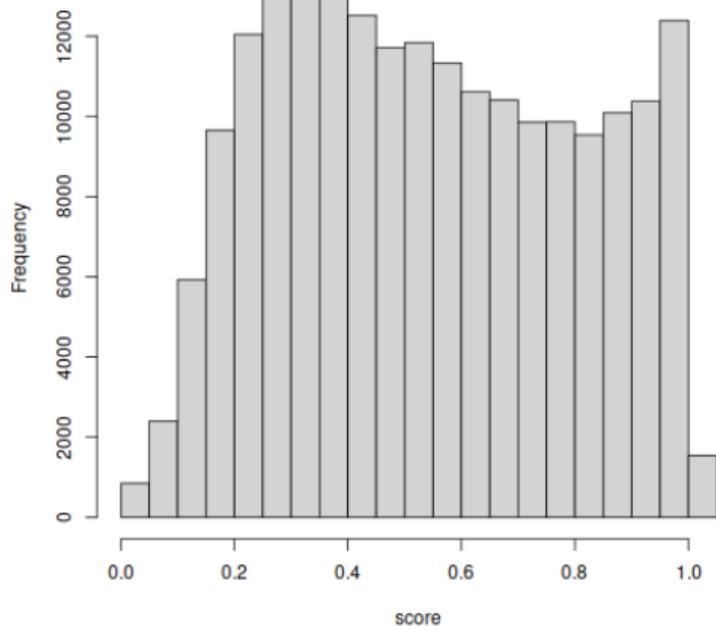
- 1 The continuum of similarity  
identical – similar – partially similar – short common passages
- 2 Similarity scores in relation to metadata  
e.g. distance in time and space
- 3 Network analysis

# THE CONTINUUM OF SIMILARITY

Similarity scores



Similarity scores to the closest neighbor



# THE CONTINUUM OF SIMILARITY

Oleks minu olemine  
Teiseks minu tegemine  
Küll ma teaksin, mis ma teaksin  
Ma teeks talli taeva alla  
Talli sisse lateri ja  
Laterisse laugu täku  
Laugu täkule sadula  
Sadulasse saksa poisi  
Saksa poisile kübara  
Kübaralle kuldapaela.

96%

Oleks minu olemine  
teiseks minu tegemine  
küll ma teaksin mis ma teaksin  
ma teeks talli taeva alla  
talli sesse lateri  
ja laterisse laugu täku  
laugu täkule sadula  
sadulale saksa poisi  
saksa poisile kübara  
kübarale kuldapaela.

Oleks minu olemine  
Teiseks minu tegemine  
Küll ma teaksin, mis ma teaksin

Ma teeks talli taeva alla  
Talli sisse lateri ja  
Laterisse laugu täku  
Laugu täkule sadula  
Sadulasse saksa poisi  
Saksa poisile kübara  
Kübaralle kuldapaela.

63%

Oleks see mu olemine  
Teiseks minu tegemine  
Ma teeks ilma ümmarguseks,  
Taevartiigi ruuduliseks,  
Ja teeks talli taeva alla,  
Talli sisse latterad,  
Latterisse laugud ruunad,  
Laugud ruunad sadulas.  
Sadulatel saksa poisid,  
Saksa poistel kübarad,  
Kübaratel kullast poordid,  
Kuldse poordil kudujad  
Kudujatel kuued teljed  
Kanguritel kaheksad teljed

Oleks minu olemine  
Teiseks minu tegemine  
Küll ma teaksin, mis ma teaksin  
Ma teeks talli taeva alla  
Talli sisse lateri ja  
Laterisse laugu täku

15%

Laugu täkule sadula  
Sadulasse saksa poisi  
Saksa poisile kübara  
Kübaralle kuldapaela.

Oleks minu olemine  
Teiseks minu tegemine  
Küll ma teaksin, mis ma teaksin  
Ma teeks talli taeva alla  
Talli sisse lateri ja  
Laterisse laugu täku  
Laugu täkule sadula  
Sadulasse saksa poisi  
Saksa poisile kübara  
Kübaralle kuldapaela.

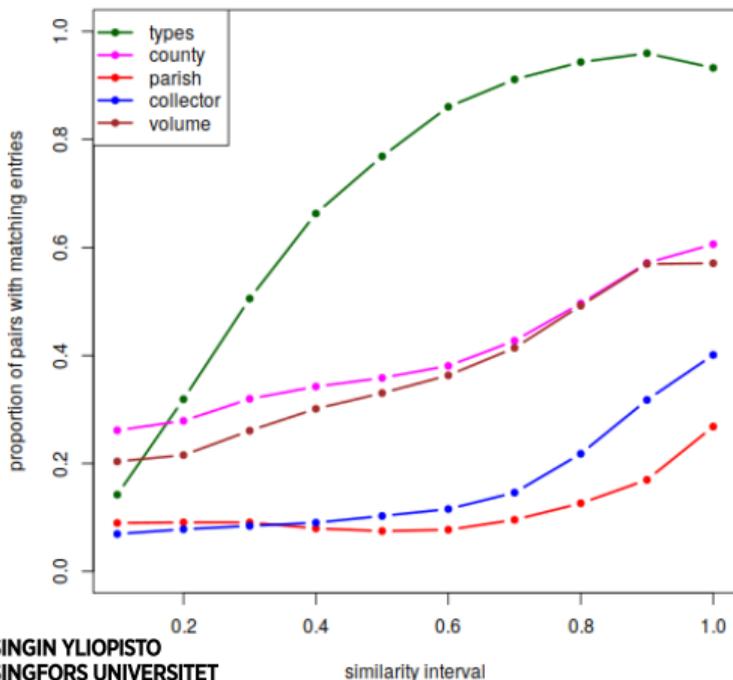
37%

Laugu täkule sadula  
Sadulasse saksa meesi  
Saksa meesile kübara  
Kübarale kuldakrooni  
Kuldakroonile kuduja  
Kudujale terav mõeka  
Kuldamõegale tegija  
Tegijale terav kerves

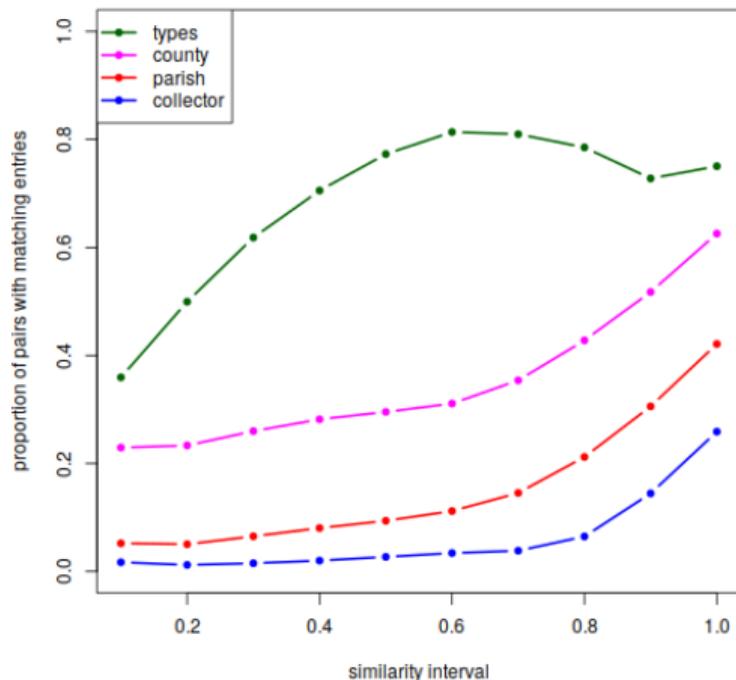
Oh seda venda mis minule  
Ja seda ratsu, mis minu ver  
Vend oli isi ilmatarka,  
Ta tegi talli taeva alla  
Talli sisse lateruse,  
Laterusse lauku täkk.  
Ilm ei jõudnud ehitada,  
Vald ei jõudnud vallid teha,  
Kihelkond kinni pidada.  
Mina vennal üttelema:  
Vennakene, elakene,  
Vala tal pähä vaskpäitsed,  
Pane tal luine looka peale  
Pane koela raudarangid,  
Mine siis soosse sõitema  
Arusse teeda ajama.  
Kui põle soosta sõidetud  
Ja arusta teed aetud:  
Mine siis läbi alta ilma,  
Alta ilma, pealta päeva,  
Viie vikerkaare vahelt  
Ja kuie kuu keske'elt!

# SIMILARITY VS. METADATA

SKVR: Matching metadata entries for pairs of similar poems

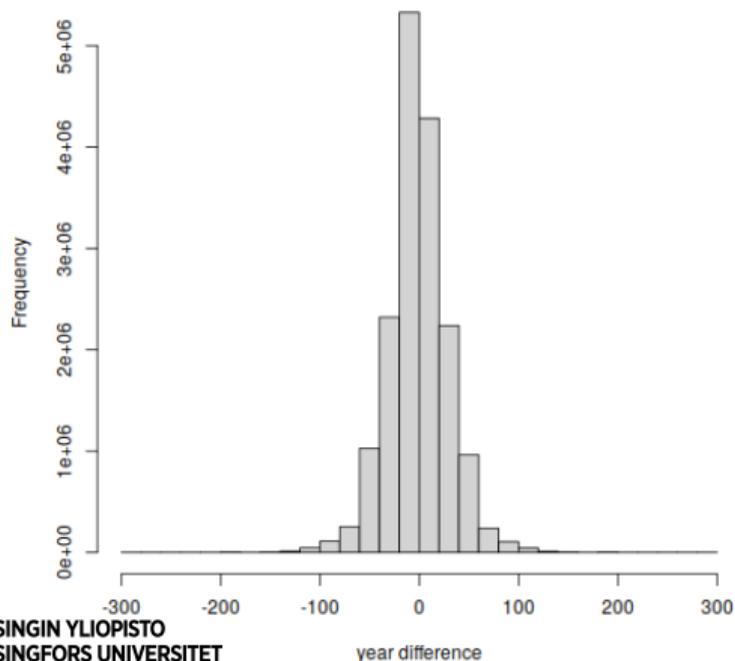


ERAB: Matching metadata entries for pairs of similar poems

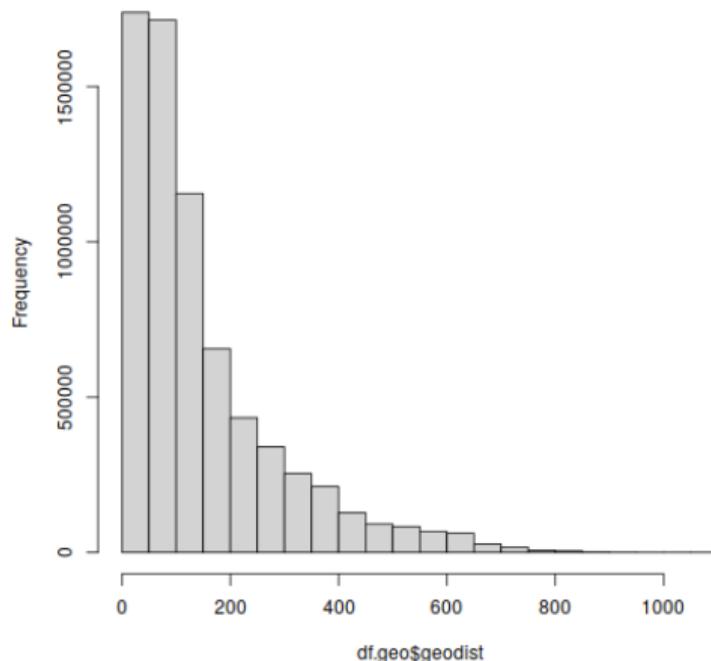


# SIMILARITY VS. METADATA

Histogram of year difference for similar poems



Histogram of distance (km) of similar poems



# OUTLIER ANALYSIS

Similar poems with the largest temporal distance:

- charms (e.g. 1 2 3 4 5)
- proverbs (e.g. 1 2 3)
- epics: 'The spider song' (1)

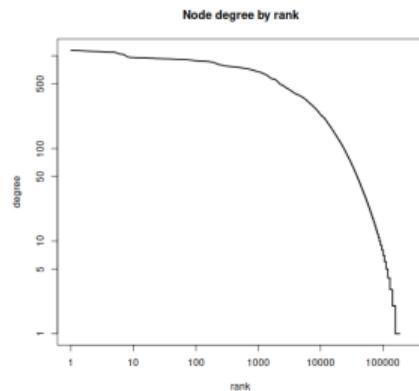
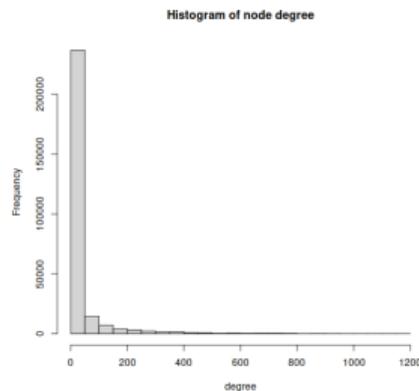
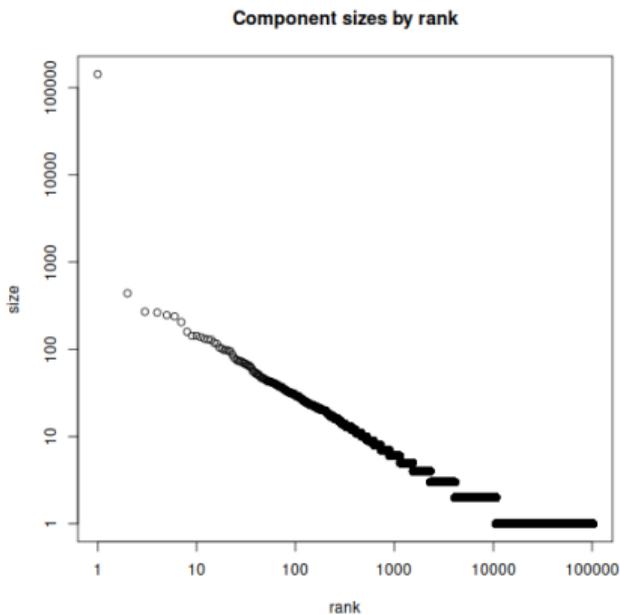
Similar poems with the largest geographic distance:

- charms (e.g. 1 2)
- lullabies (e.g. 1 2)
- children's songs (e.g. 1 2)

# NETWORK ANALYSIS

Basic facts:

- high clustering coefficient (0.76)
- very skewed degree distribution
- The **giant component** contains:  
58% ERAB, 43% JR, 53% SKVR

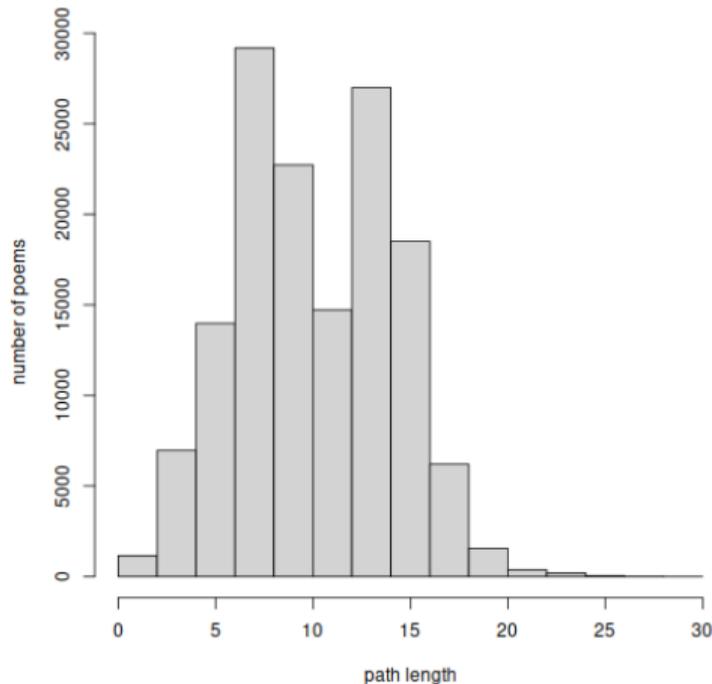


# EXAMPLE: IRON CHARM

Voipa sinua rauta raukka,  
Etpä silloin suuri ollut,  
Etpä suuri, etpä pieni,  
Etpä kovin korea,  
5 Kun sinua suosta sotkettihin,  
(Vetelästä vellottihin),  
Kun sinä maitona (makasit)  
Nuoren (neitosen nisissä,  
Kasvannaisen kainalossa).

*SKVR XIII 4112.*

Number of poems reached from skvr12141120 by path length



# iron charm → snake charm → short epic motif → lullaby

Voipa sinua rauta raukka,

Etpä silloin suuri ollut,  
Etpä suuri, etpä pieni,  
Etpä kovin korea,  
5 Kun sinua suosta sotkettihin,  
(Vetelästä vellottihin),

Kun sinä maitona (makasit)  
Nuoren (neitosen nisissä,  
Kasvannaisen kainalossa).

Mato musta maan alanen,  
Toukka tuomen karvallinen,  
Joka läpi mättäihien mänöö,  
Puun juuret pujotteloo.  
5 Et ollut silloin suuri,  
Et suuri etkä pieni,

Kun suosta sotkuttelin,  
Vetelästä vellottelin.  
Et ollut silloin suuri,  
10 Etkä suuri etkä pieni  
Kun maassa maitona makasit,  
Nuoren neitosen nisässä,  
Kasvannaisen kainalossa.

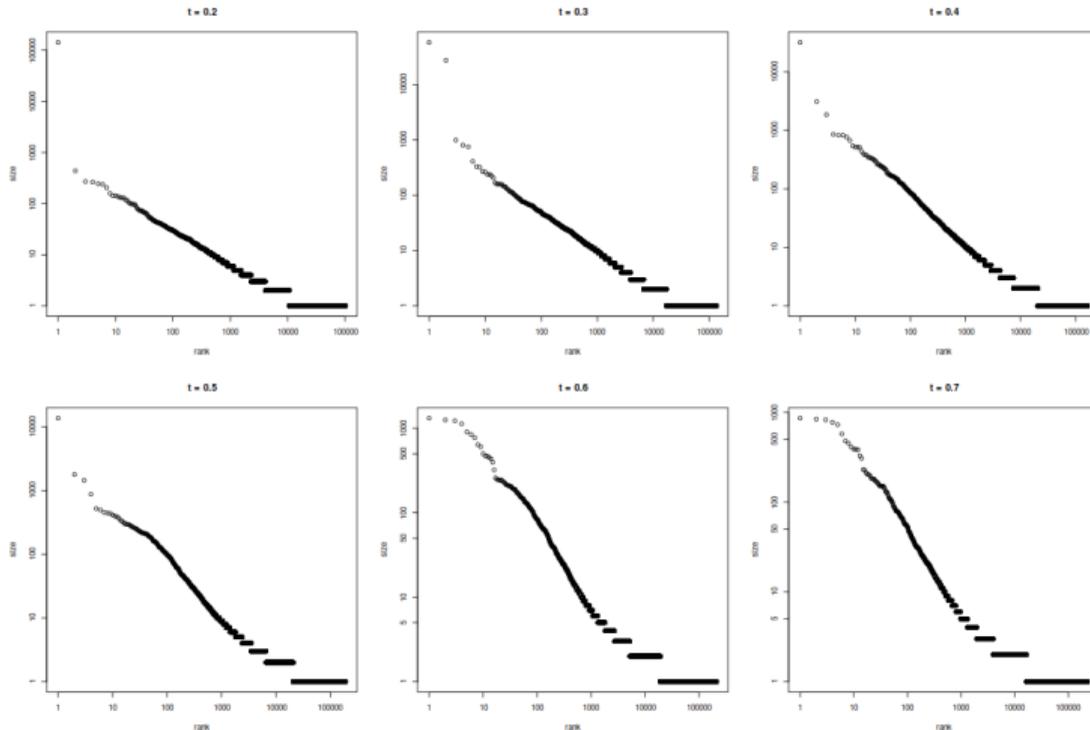
Mansikka mäellä huusi:

"Tule, neito, noppimaan,  
Punaposki poimimaan  
Ennenkun etana syöp',  
5 Musta mato mustentaa,  
(Ennenkun etana syöp',  
Mato musta maan alainen,  
Toukka tuomen karvainen.)"

Nuku nurmilintu,  
Väsy, väsy västäräkki,  
Nuku, kun mie nukutan,  
Väsy, kun mie väsyttän.  
5 Mantsikka mäellä huus,  
Puola kaiku kankahalla:  
Tule neito noppimaha,  
Punaposki poimimaan,  
Ennenkuin etana syöpi,  
10 Mato musta muihkajaapi.

Tule uni ukko meille,  
Tuo unta tullessaisi  
Kokonainen kontillin.  
Uni ulkona kysyvi,  
15 Unen poika porstuasta:  
Onkos lasta laattialla,  
Vakahaista vaattehissa,  
Pientä peittehen sisällä?

# COMPONENT SIZES



# SKVR-ERAB “BRIDGES”

## SKVR III3 3531.

skvr03135310  
Länsi-Inkeri — Narvusi  
1909 Lukkarinen, Juho

Tups, tups, tups, Martikkine,  
Siiher saaher Santikkine,  
Ei ole Martti maasta tullut,  
Martti on tullut taivahasta  
5\_Kultasta korentaa myö(ten),  
Hopiaista ortta myö(ten),  
Vaskista varoa myö(ten).  
Tuokaa Martille munia#1,  
Saakaa Santille papuja.

## EÜS X 1246/7 (2) <

ey1012460002  
Harjumaa — Kuusalu  
1913 Raudkats, Anna

Uups, uups, Mardikane,  
Seer, seer, sandikane  
Ei ole Marti maasta tulnud  
Mart on tulnud taeva'asta  
Kuldasta korenda mööda  
Hobeasta orta mööda  
Vaskista vabeta mööda  
  
Tinasta teerada mööda.

(10 steps to the iron charm)

# NETWORK ANALYSIS: CONCLUSIONS

The giant component arises due to:

- mixing of motifs
- reuse of common formulas in different contexts

The network structure is thus **characteristic of oral poetry** and unusual for text similarity in general.