

Geography of Russian poetry: Countries and cities inside the poetic world

B. Orekhov, E. Kuzmenko
nevmenandr@gmail.com, lizaku77@gmail.com

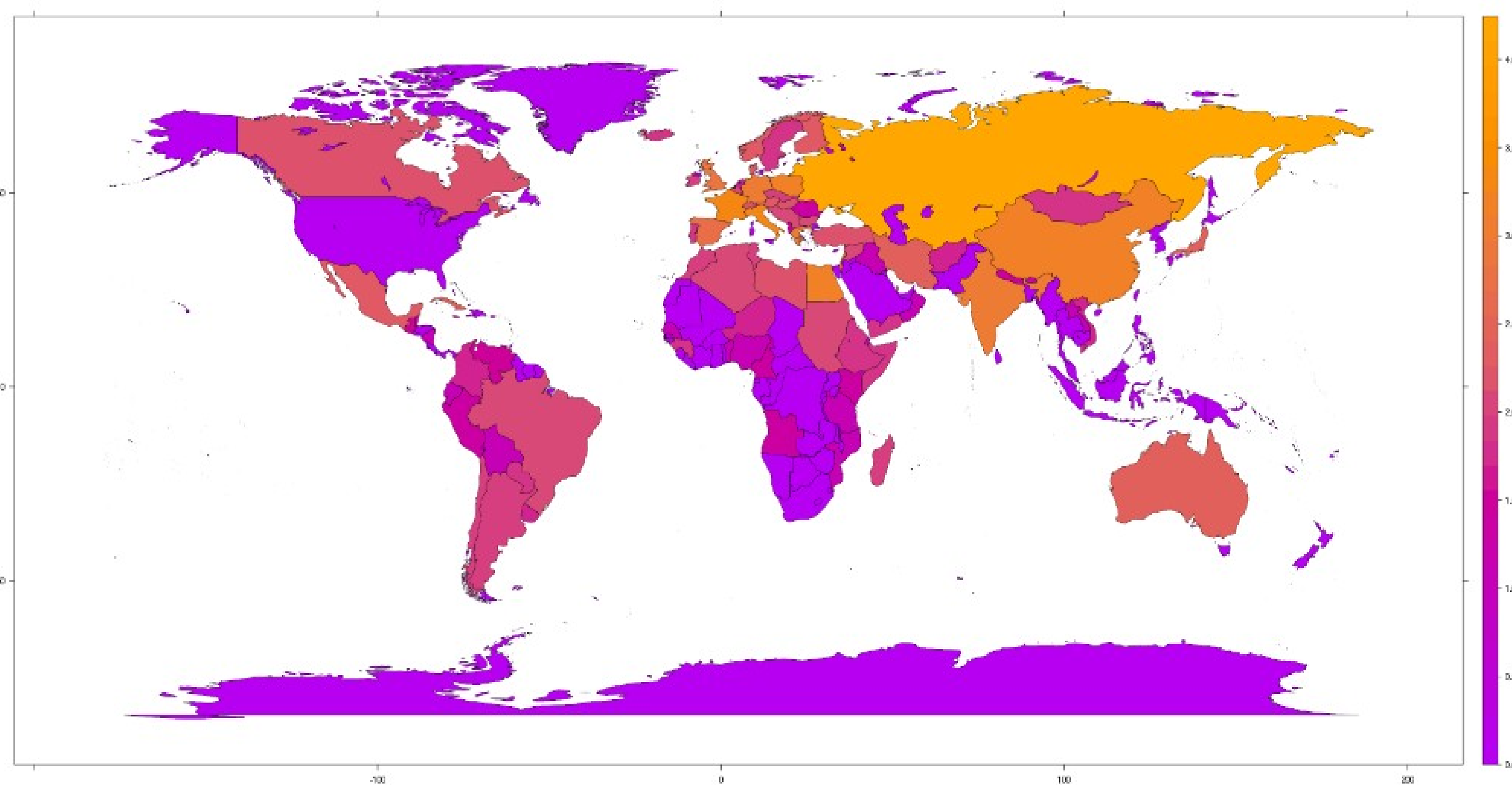
Our goals:

- extract names of countries and cities from poetic texts;
- reveal major trends in functioning of extracted entities;
- link these trends to poetic traditions;
- analyze linguistic features of toponyms used in poetry.

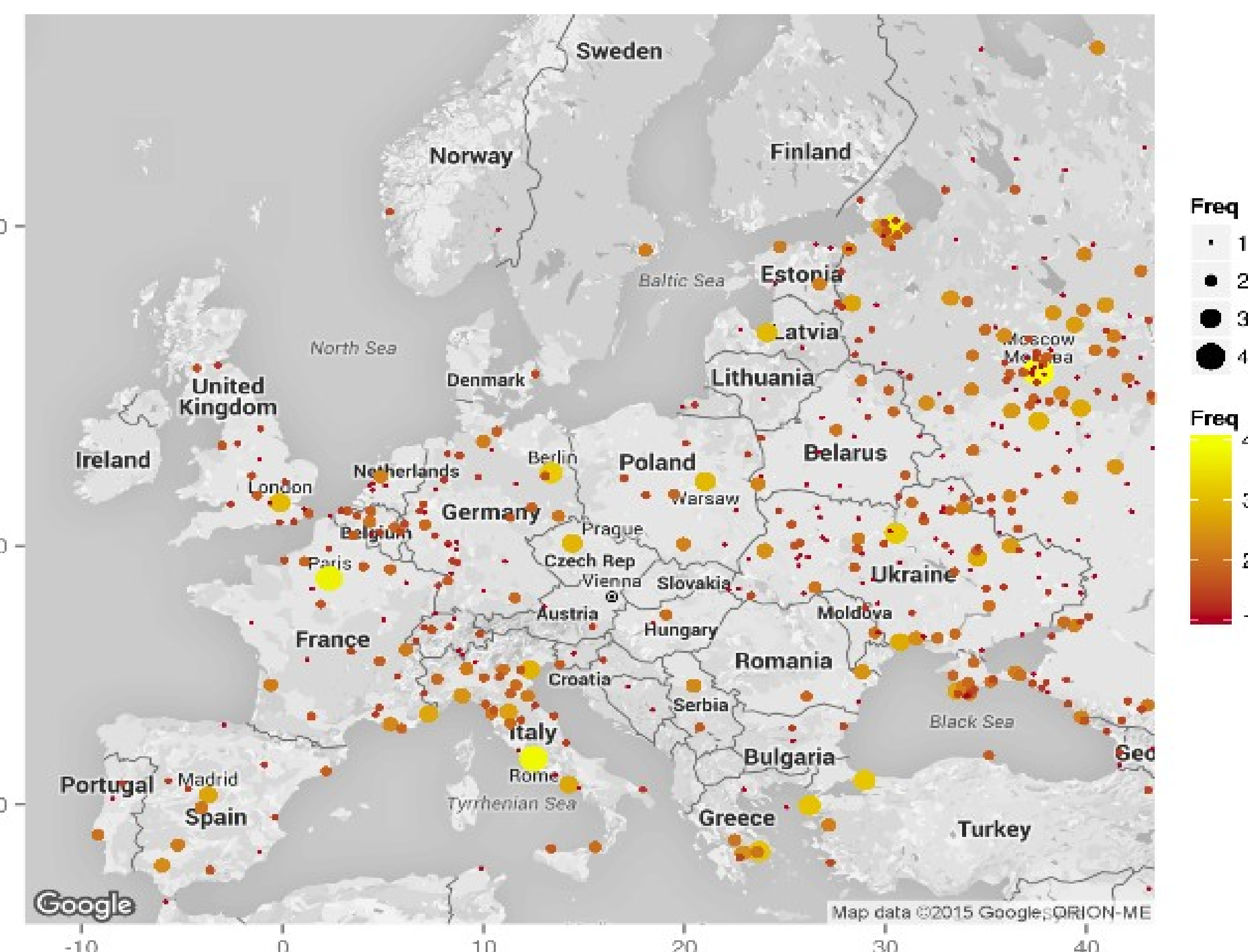
Tools and methods:

- poetic subcorpus of the Russian National Corpus;
- list of toponyms from Wikipedia;
- graph analysis and visual analysis.

Frequencies of country and city names in the whole corpus

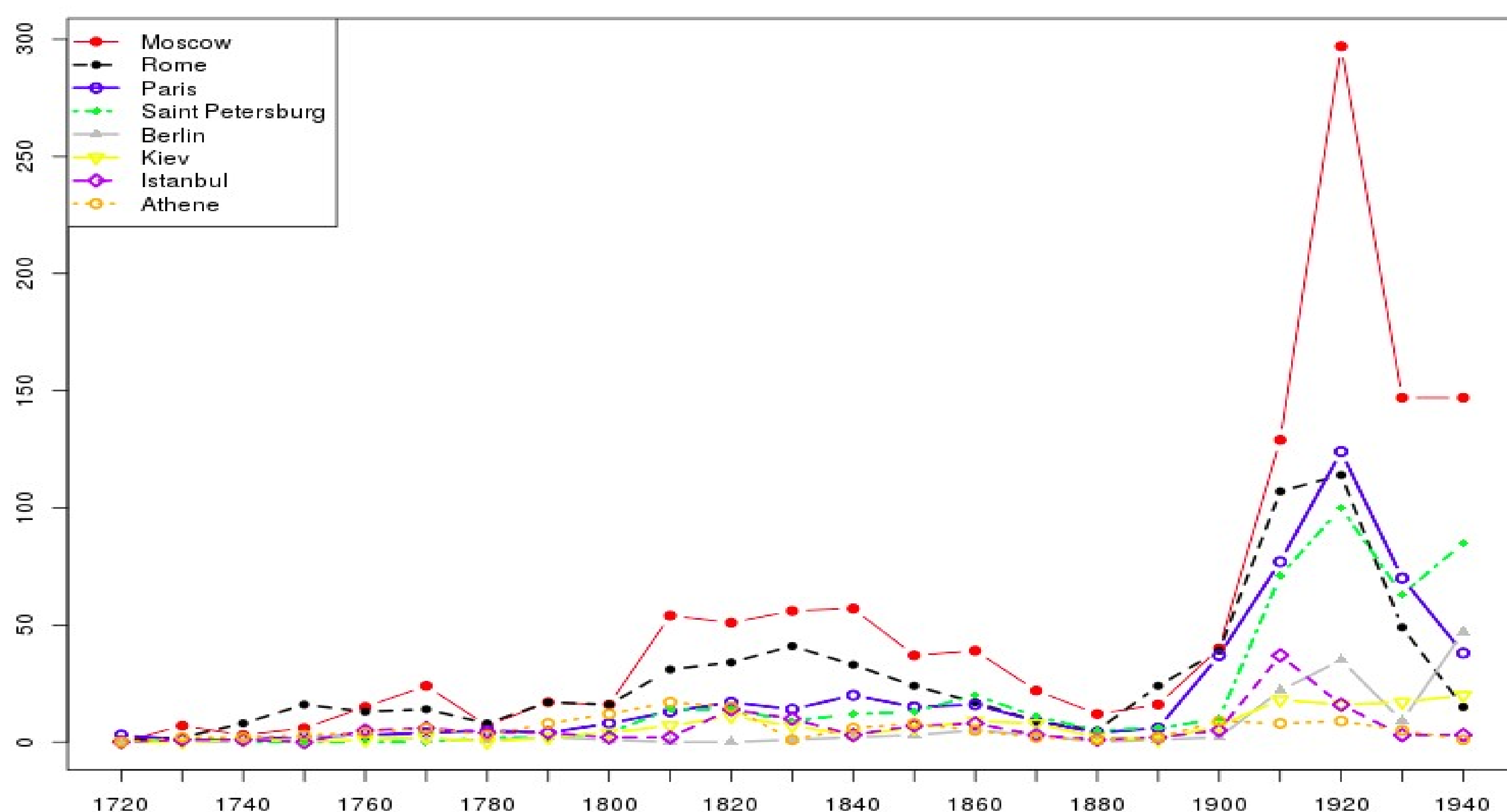


Russia (obviously), China, Egypt, Romanic European countries and specific Russian East (Georgia, Iran) dominate in the poetic world.



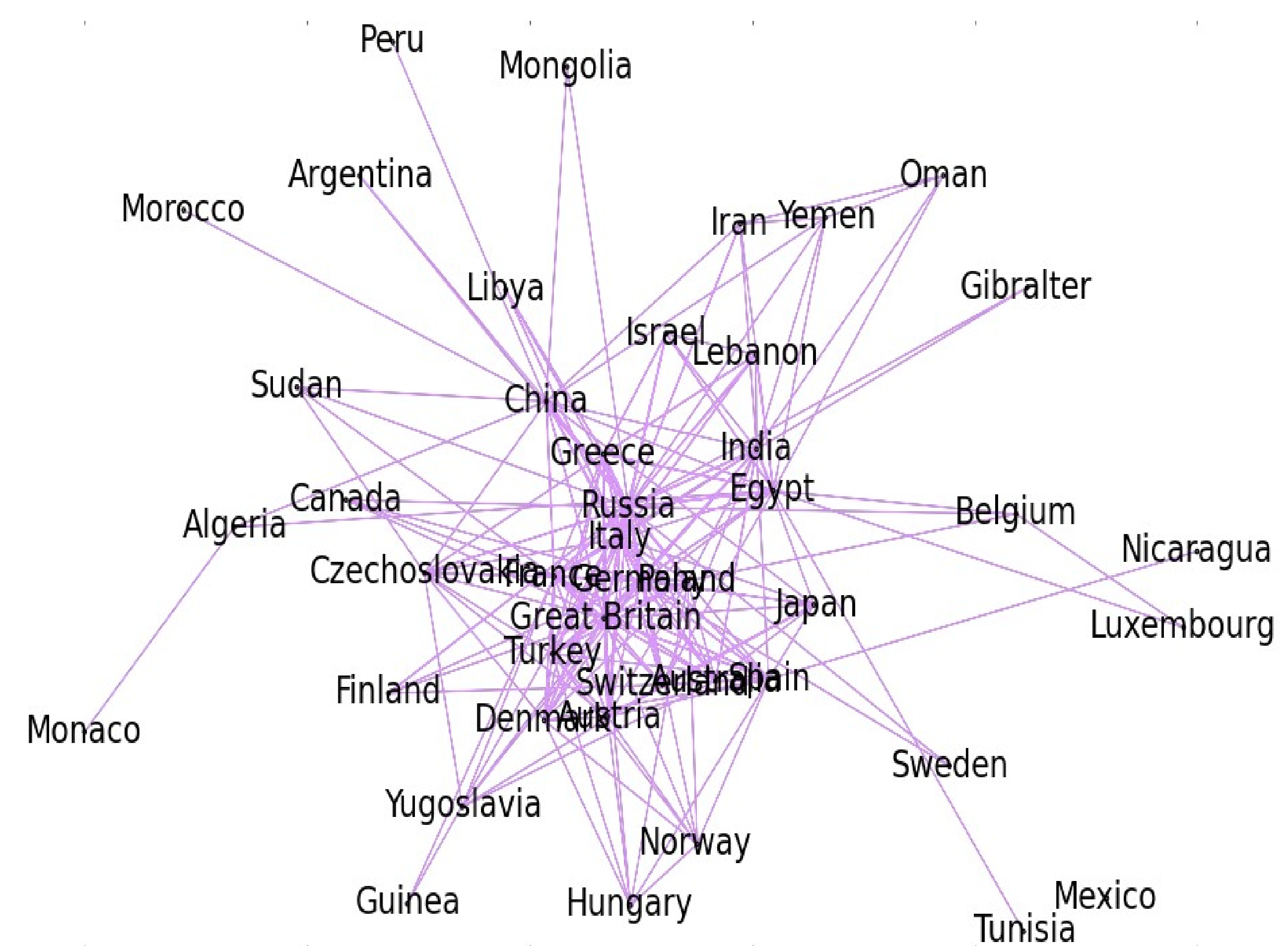
Areas near Moscow, Crimea and Mediterranean seaside are most charged with poetic cities.

Temporal tendencies for city names



The burst of usage can be found in the beginning of the 20th century. At the same moment numbers begin to differ greatly.

Graph of co-occurrences for country names



If two country names are met together in a poem, they are connected in the graph. The nodes with the highest centrality score are Russia, Italy, China and Egypt.

Conclusions

- Not all cities and countries are alike in poetry.
- The distribution of country and city names is linked to certain poetic traditions and has its dynamics.
- Interest to geography weakens in the 80s both in XIX and XX centuries and strengthens in the beginning of every century.

Problems

- Entity recognition systems work pretty bad for poetic texts.
- Homonymy of geographic names mixes things up: Togo (pronoun), Sofia (first name), Belmez (idiom)