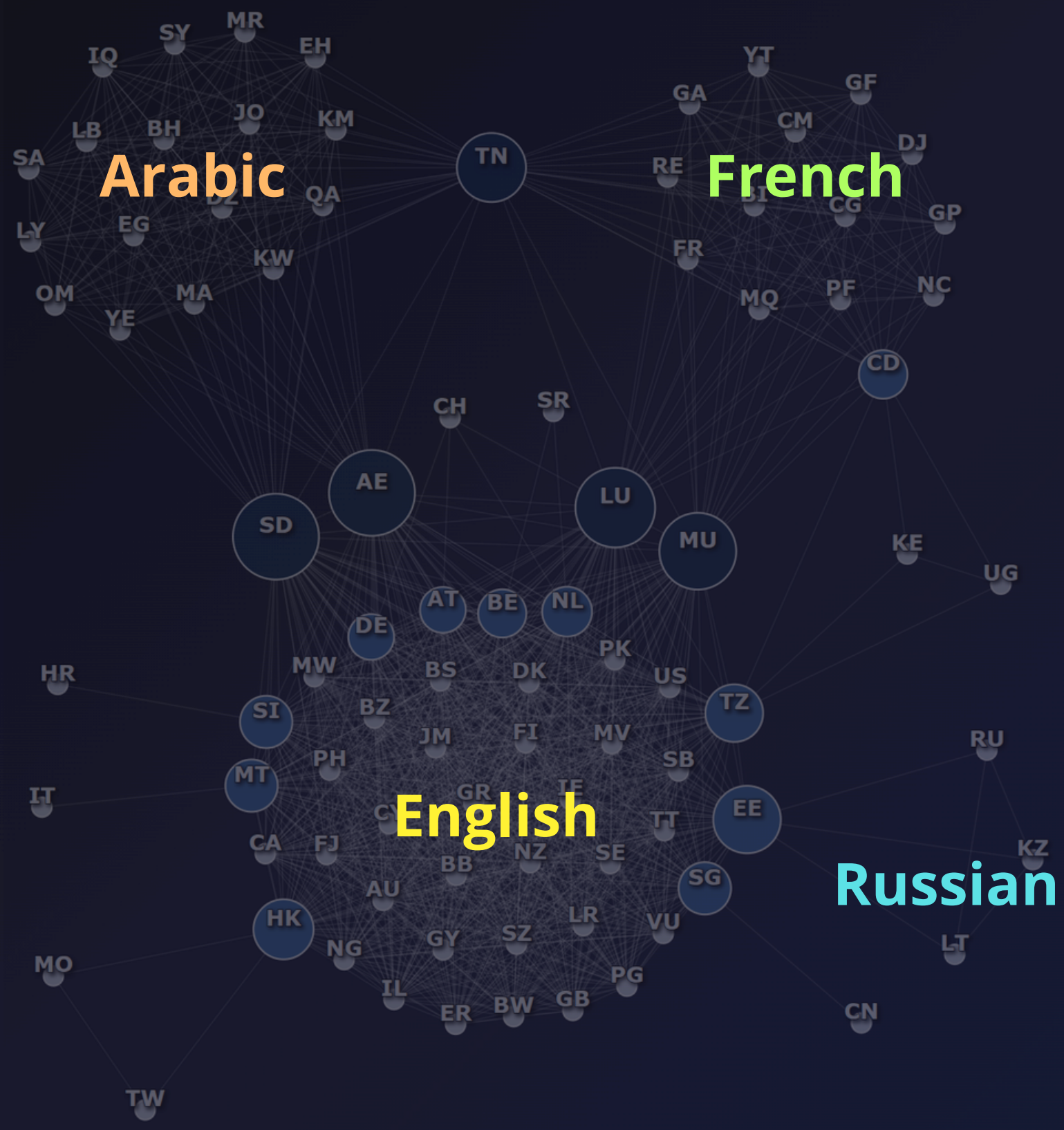
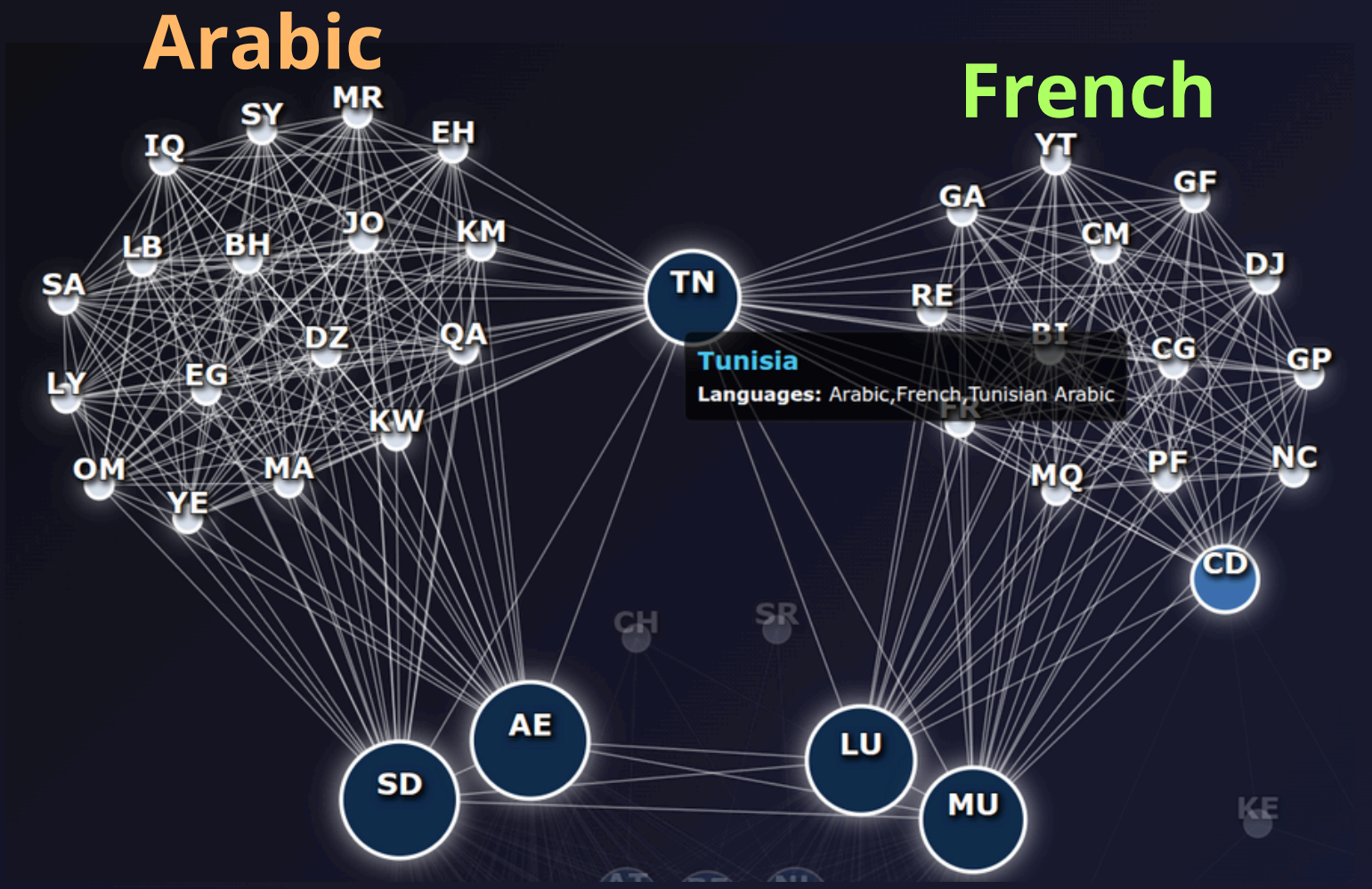
Where Languages Meet Worlds

How Unicode reveals the hubs and bridges that shape global linguistic connectivity



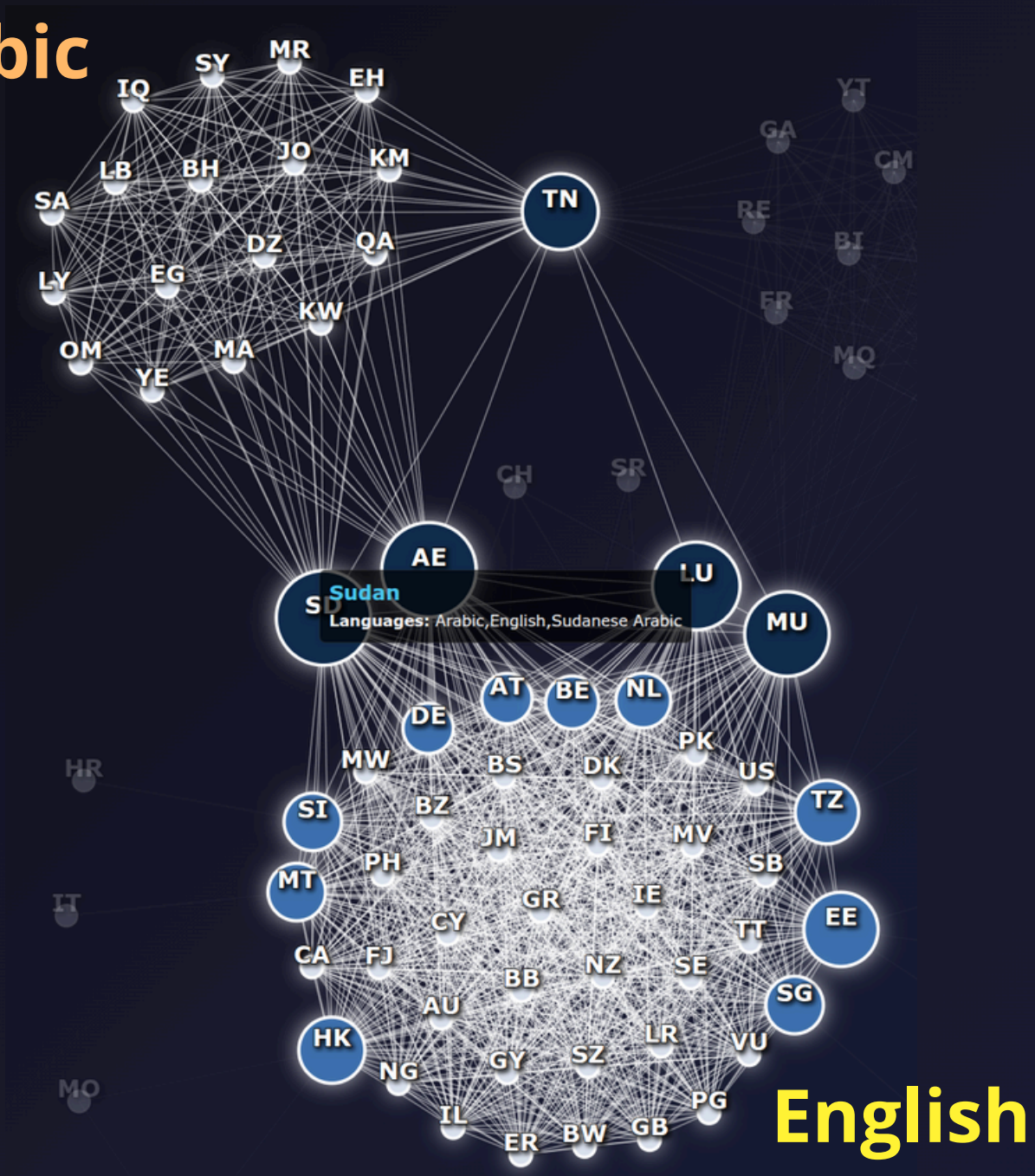
- Unicode was created to standardize the world's writing systems.
- But it also provides language–population data, showing what share of people in each country speak each language.
- I built a country-to-country network: nodes are countries (pop. ≥200k), edges connect when share a language (≥50%).

Tunisia:
Arabic & French



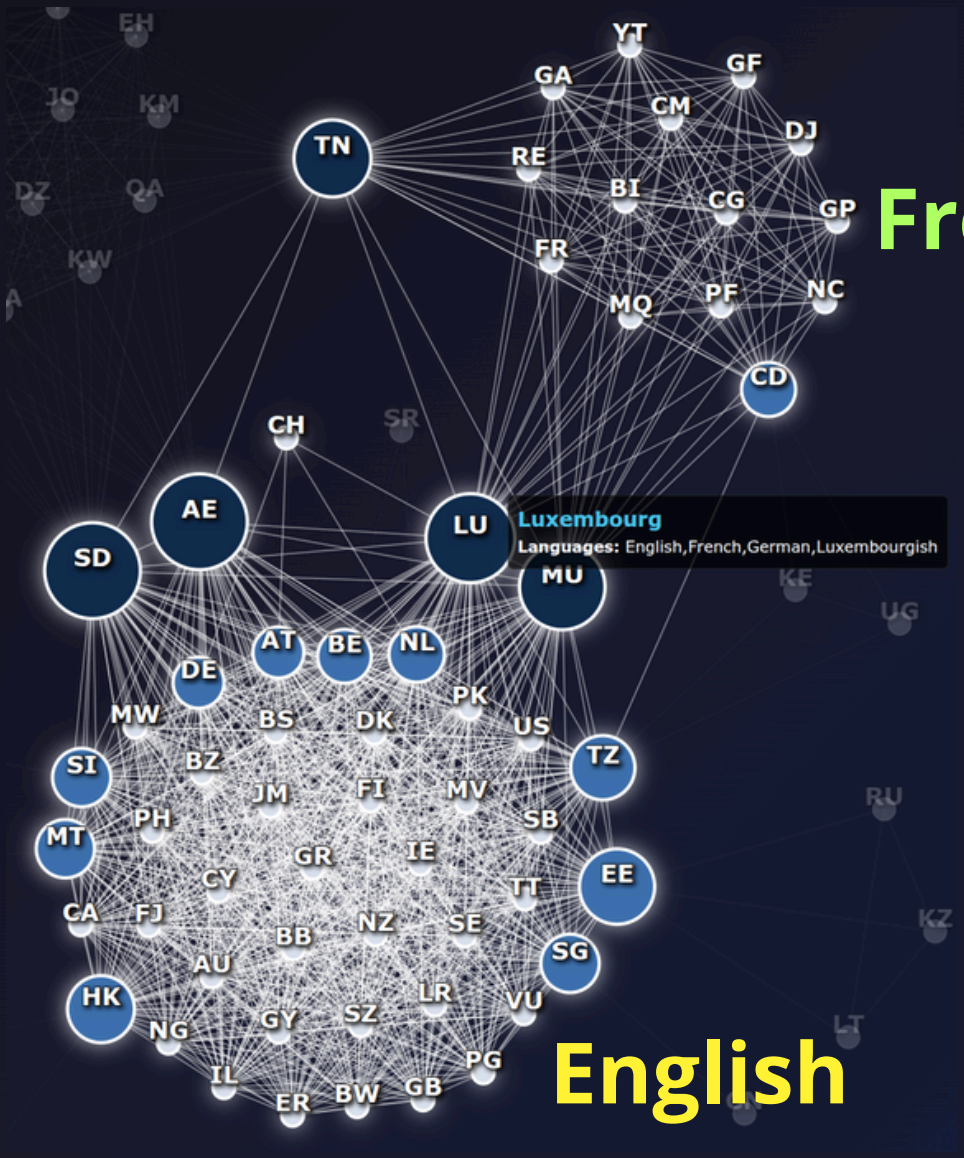
Tunisia (90% Arabic, 53% French) is perhaps the most important node in this network, serving as the sole bridge between the Arabic cluster and the Francophone cluster.

Sudan and United Arab Emirates: Arabic & English



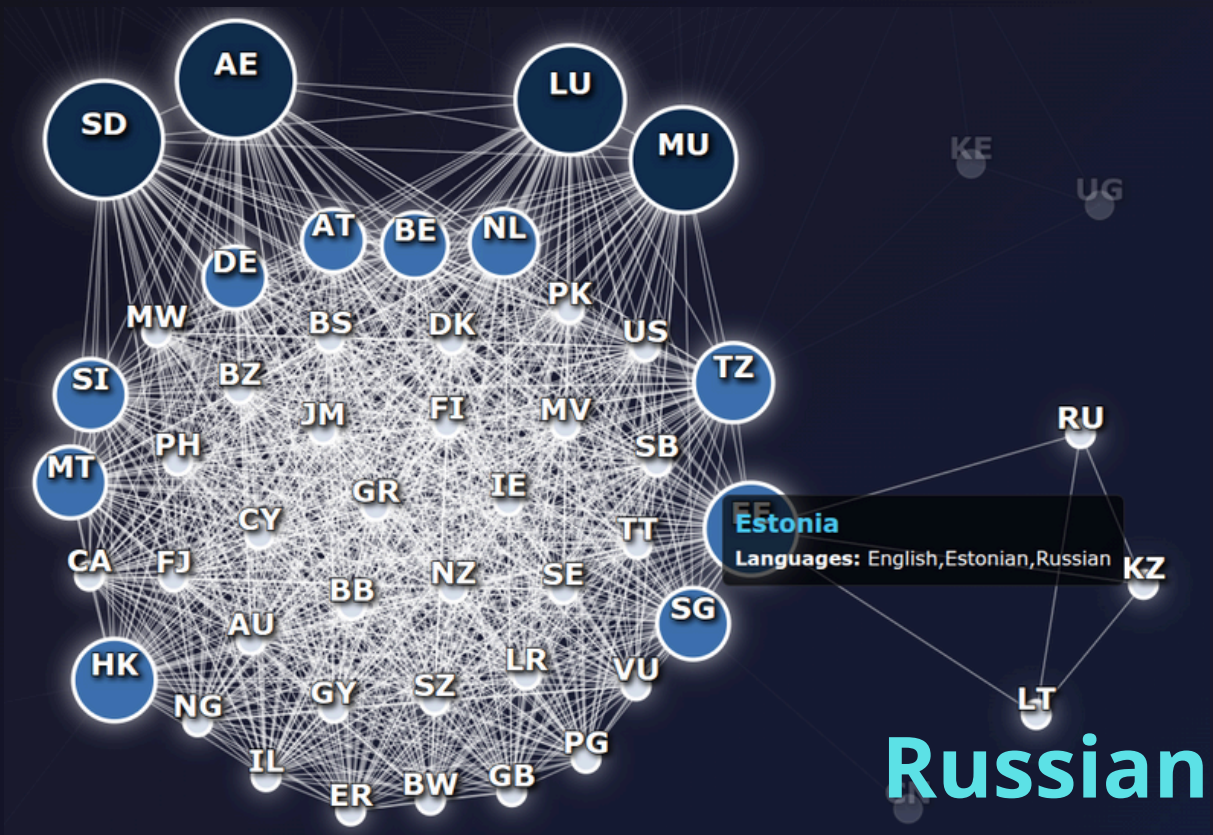
In Sudan (61% Arabic, 61% English) and the UAE (78% Arabic, 50% English), bilingualism creates strong internal bridges, placing them at the crossroads of Arabic and Anglophone clusters.

Luxembourg & Mauritius: English & French



Luxembourg (92% French, 56% English, 63% German) and Mauritius (73% French, 72% English) embody multilingual overlap, tying the French and English clusters together while Luxembourg adds a German link within Europe.

Estonia:
English & Russia



Estonia (56% Russian, 50% English) links the global Anglophone core with the Russian-speaking cluster of Russia, Latvia, and Kazakhstan.

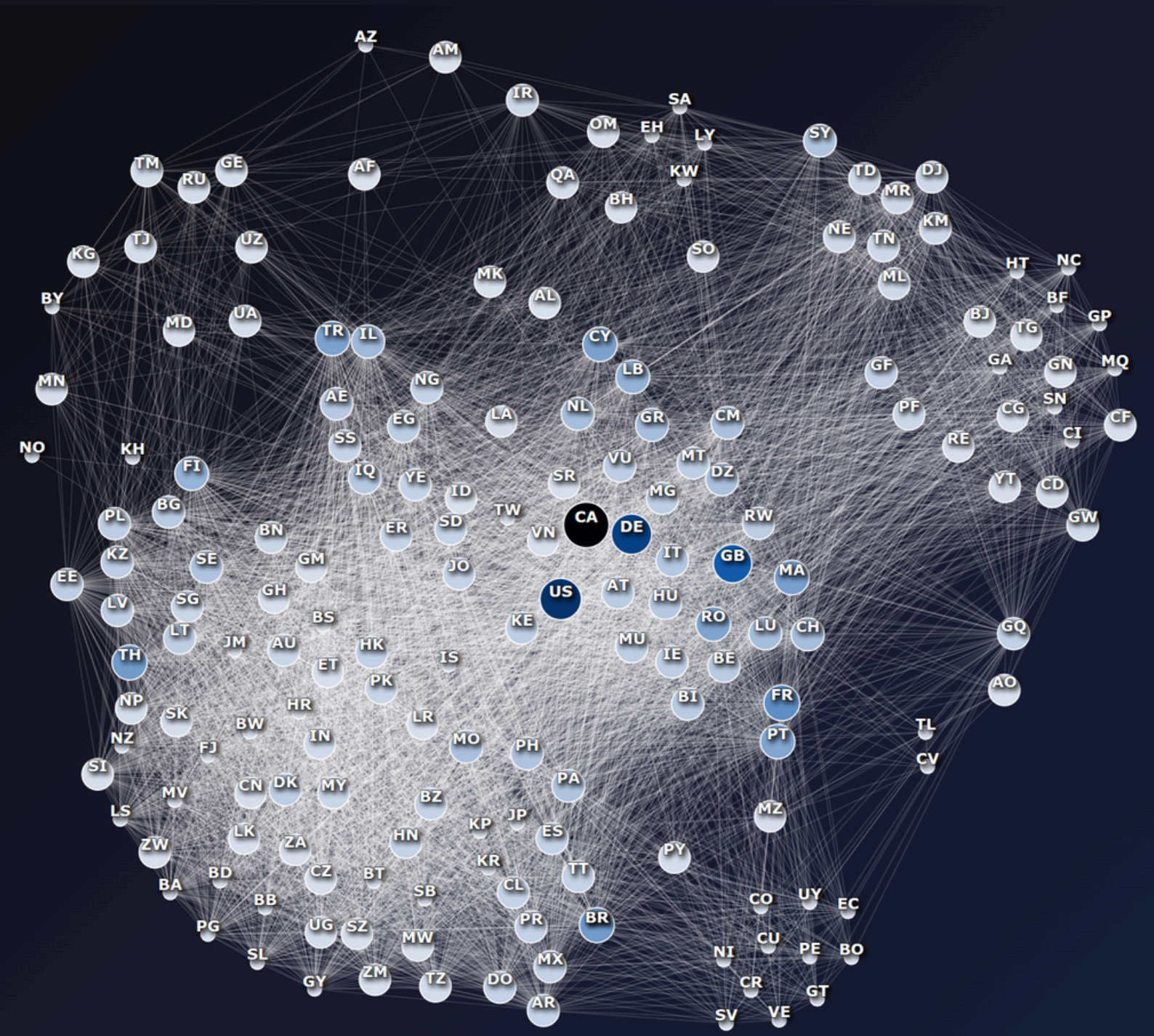
Data sources:

- https://www.unicode.org/cldr/charts/47/supplemental/territory_language_information.html (2025 updated)
- World Bank Python API (https://pypi.org/project/wbgapi/)
- https://networks.skewed.de/net/unicodelang (2015)

Check it out more on:

alvarofrancomartins.com/post/language-network

For reference, here's the network when we lower the language-speaking threshold to ≥ 0.01% of the population.



The world is much more connected than we think.