

INTRODUCTION

Cultural diaspora and mobility arise from political change, institutional opportunity, and patterns of human movement. People relocate for jobs, in response to economic shifts, or under environmental pressure. Mobility is also driven by creative aims, as artists learn from new contexts while contributing to them in an ongoing exchange of ideas.

We took ballet, a classical European form now practiced worldwide, as a lens on how creative work travels. Using English-language Wikipedia as a large, publicly maintained catalogue of ballet works and artists, we extracted relatively consistent records of premieres and contributors across countries, while remaining mindful of its biases. With LLM-based metadata extraction, we turned these dispersed entries into structured data for scalable, reproducible study of cultural mobility in the performing arts.

DATA

- Extracted from the English Wikipedia page *List of ballets by title*.
- Each listed title corresponds to a distinct ballet work with its own Wikipedia entry.
- Total collected entries: 518 ballet articles.

METHOD

- Applied GPT-3.5 and GPT-4 models to extract structured information (*premiere year, location, and creative contributors*) from 518 Wikipedia articles using customized prompt templates.
- Structured data verified 25% of entries for accuracy.
- Constructed a bipartite network linking artists ↔ cities to map patterns of cultural mobility and identify creative hubs.

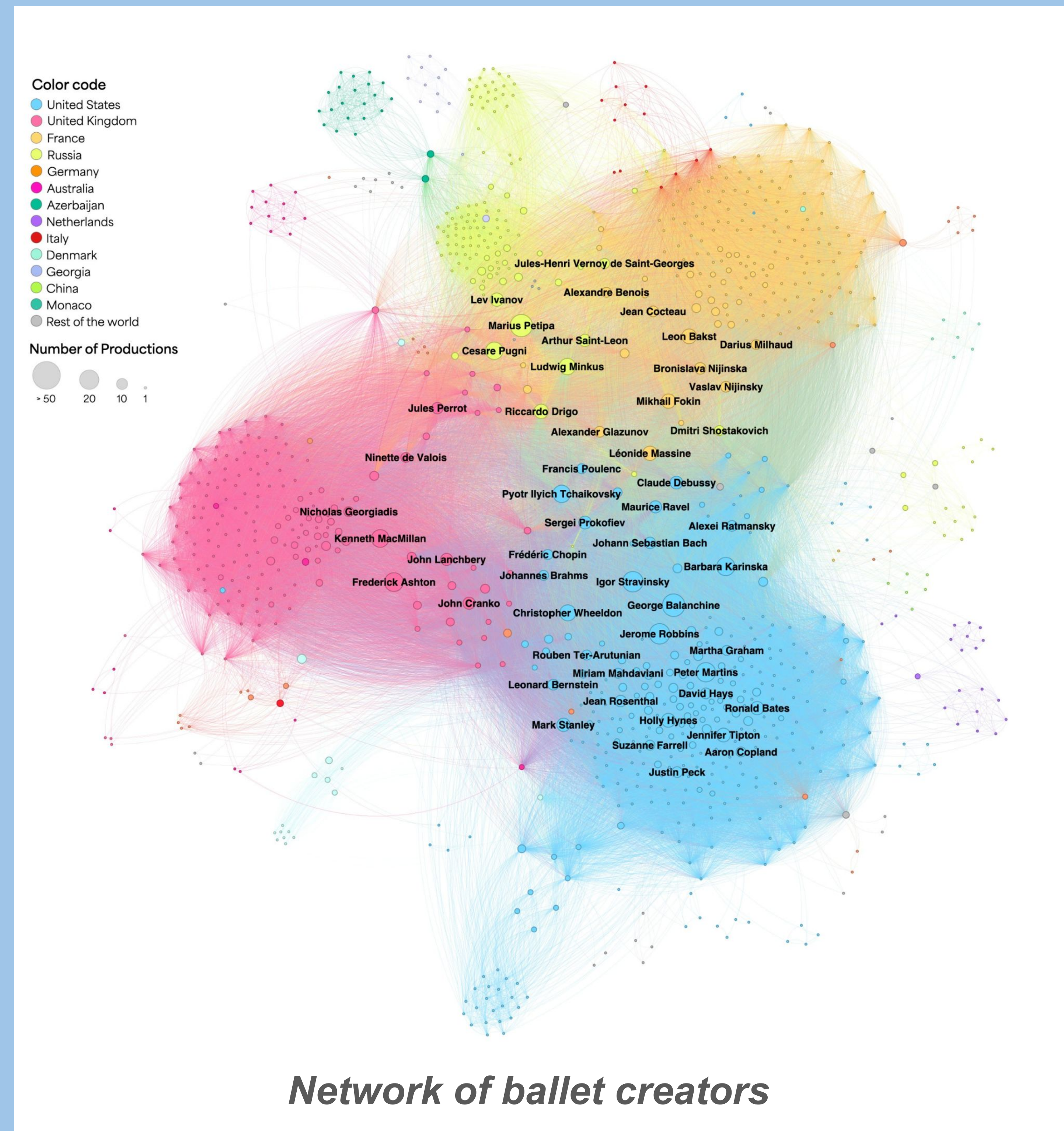
Example

Wikipedia Excerpt

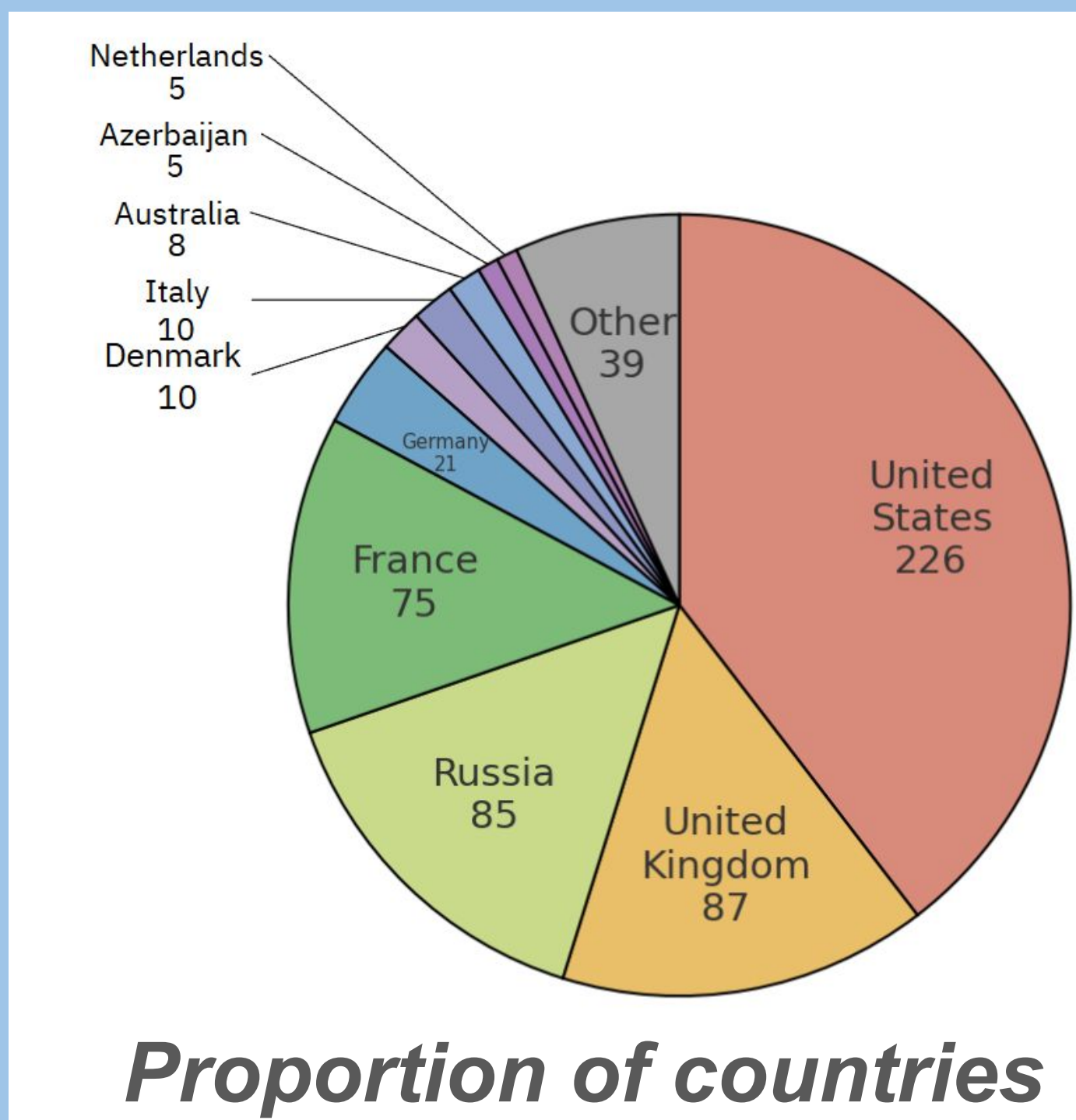
"... A Folk Tale (Danish: Et Folkesagn) is a ballet in three acts, created in 1854 for the Royal Danish Ballet by the Danish ballet master and choreographer August Bournonville to the music of Johan Peter Emilius Hartmann and Niels W. Gade. The first performance took place on 28 March 1854. Set in the Middle Ages, the ballet tells the story of a changeling living among the trolls and elves. ..."

Structured Output

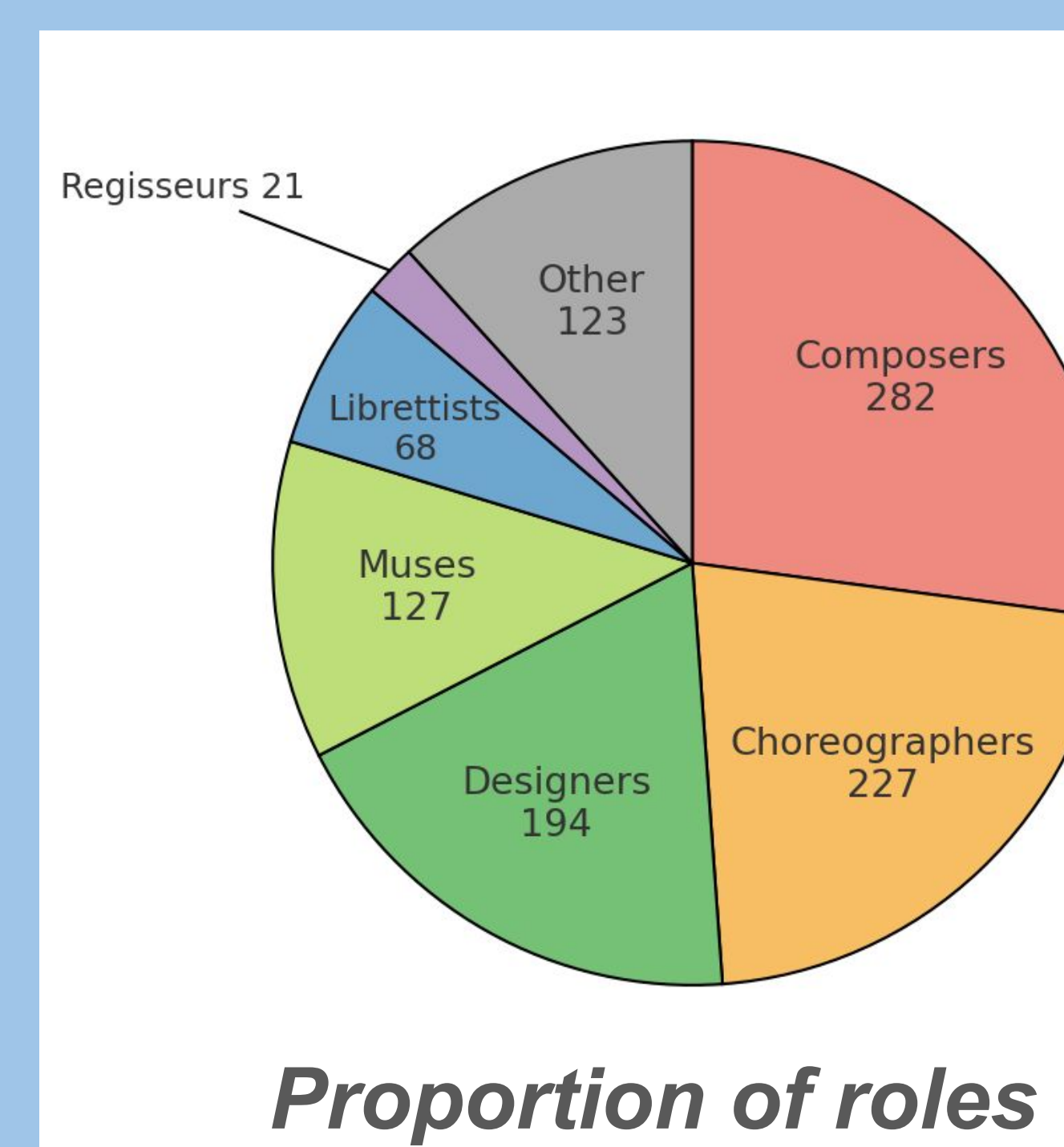
```
{
  "id": "001",
  "ballet": "A Folk Tale (Danish: Et Folkesagn)",
  "year": 1854,
  "venue": "Royal Danish Theatre, Copenhagen, Denmark",
  "choreographer": "August Bournonville",
  "composer": "[Peter Emilius Hartmann, Niels W. Gade]"
}
```



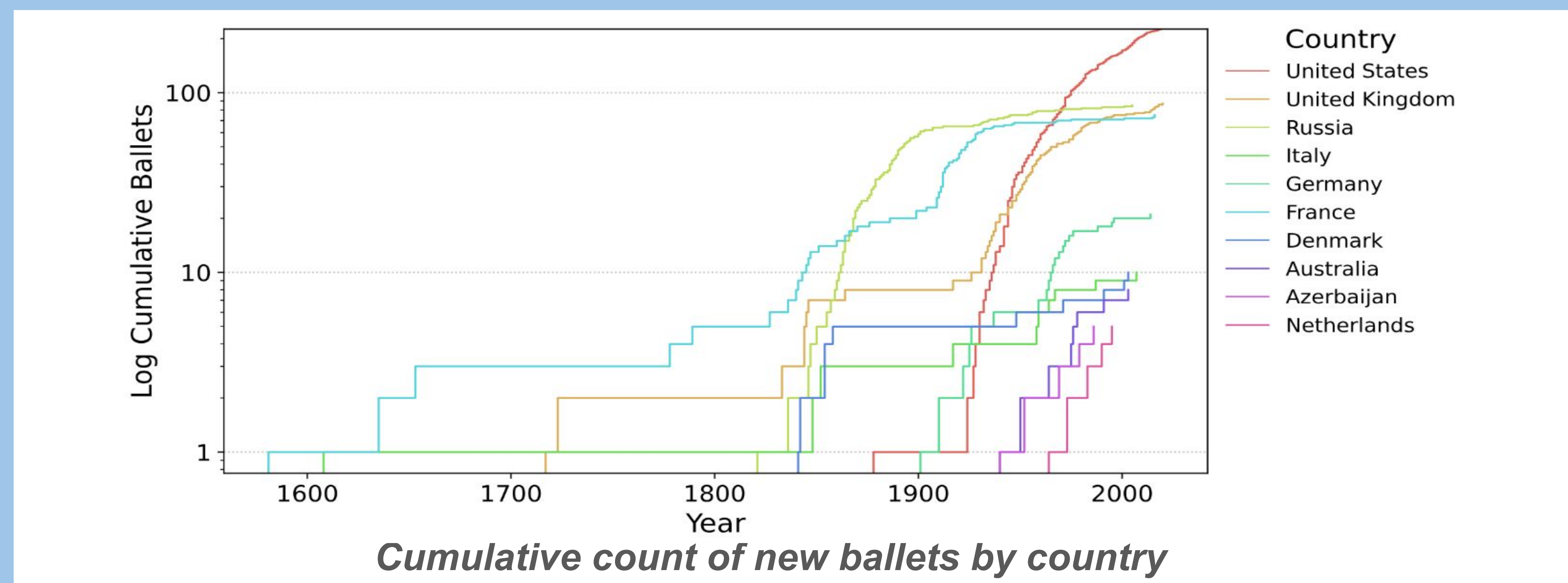
Network of ballet creators



Proportion of countries



Proportion of roles



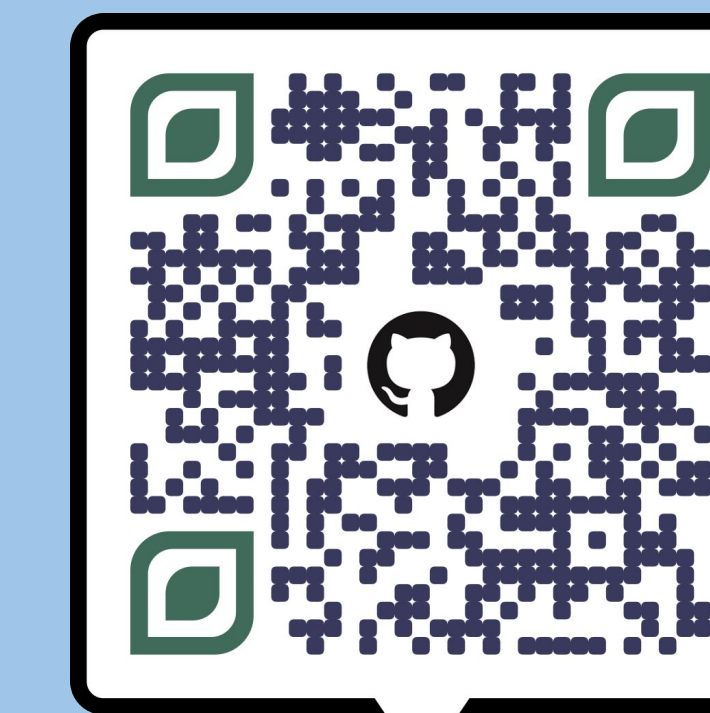
Cumulative count of new ballets by country

MODEL PERFORMANCE

Model	Accuracy	Note
GPT-3.5	0.81	Higher precision for contributor roles
GPT-4	0.72	More verbose, slightly less structured



PDF



GitHub

FINDINGS

Highly connected network: 99.88% of artists belong to one large connected component.

Key connectors: A few creative leaders bridge creative communities across countries.

Geographic patterns: Creative activity centers in the U.S., U.K., France, and Russia, with many artists producing work abroad.

Transnational flow: Ballet's evolution reflects migration and cross-cultural exchange -- core features of cultural diaspora.

Gender imbalance: Women represent only about 12% of the dataset, indicating persistent disparities in both ballet and digital records.

Temporal shift in centers: Early networks are anchored in European capitals, but from the mid-20th century onward U.S. cities become key hubs, indicating that the U.S. gradually gains the ballet market.

Wikipedia as data source: Using English Wikipedia makes this mapping feasible at scale, but it also reproduces its biases, over-representing Western institutions and well-documented artists while under-representing others.

CONCLUSION

Using Wikipedia integrated with LLMs, we convert narrative entries into structured records which can be used for network and temporal analysis of cultural diaspora. This workflow makes it possible to map creative mobility at a large scale and to link artists, places, and premieres in ways that would be difficult to do by human annotators. At the same time, it carries over the gaps and biases in the original sources, and still depends on carefully designed prompts and human checking to catch potential extraction errors.

REFERENCES

- Adamson, F. (2012). *Constructing the diaspora: Diaspora identity politics and transnational social movements*. In *Politics from afar: Transnational diasporas and networks* (pp. 25–42). Columbia University Press.
- Dzionek-Kozłowska, J., & Matera, R. (2021). Institutions without culture: On Daron Acemoglu and James Robinson's theory of economic development. *Journal of Economic Issues*, 55(3), 656–676. <https://doi.org/10.3386/w28832>
- Sorokin, P. A. (2019). *Social and cultural mobility*. In *Social stratification, class, race, and gender in sociological perspective* (pp. 303–308). Routledge.
- Czaika, M., & Reinprecht, C. (2022). *Migration drivers: Why do people migrate*. In *Introduction to migration studies: An interactive guide to the literatures on migration and diversity* (pp. 49–82). Springer International Publishing.
- Liu, L., Dehmamy, N., Chown, J., Giles, C. L., & Wang, D. (2021). Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-020-20674-1>
- Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., & Helbing, D. (2014). A network framework of cultural history. *Science*, 345(6196), 558–562. <https://doi.org/10.1126/science.1240064>