wikipedia and ai
for mapping cultural diaspora

Y. Herrera-Guzmán,[1]
Y. Peng,[2]
M. Lamba[3]

[1] University of Notre Dame
[2] University of Illinois
[3] University of Oklahoma

NetSci Boston 2026

UNIVERSITY OF NOTRE DAME

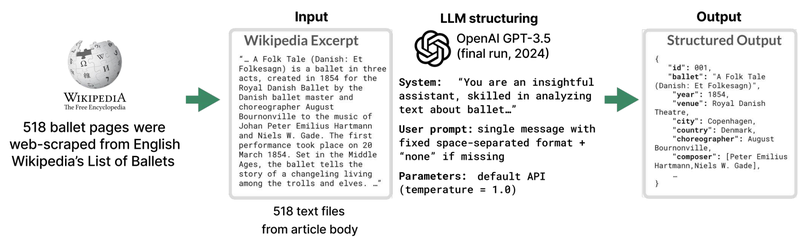
Art travels with people.
When artists migrate, they carry their craft across borders, transforming local traditions into global forms. Cultural diaspora captures this process, tracing how human mobility shapes the spread and evolution of cultural expressions across time and place.

Ballet is one of the most striking examples, a European elite art that became a worldwide cultural force through centuries of migration, exchange, and reinvention.

DATA AND METHODS

We use LLMs to convert narrative entries from Wikipedia into structured records. Our goal is to test if Wikipedia can be reliably used to extract structured information about artistic mobility, as a starting point for the systematic study of cultural diaspora.

We scrape 518 ballet entries from Wikipedia's List of Ballets by Title, then we use GPT-3.5 and GPT-4 to extract named artists, creative roles, and premiere cities from each entry.

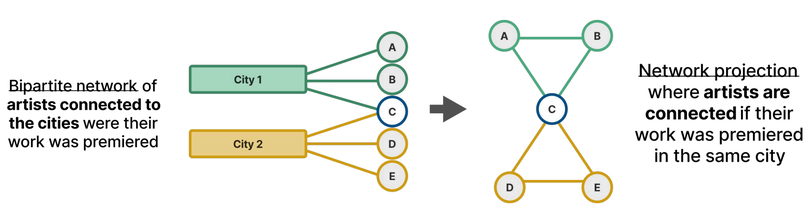


The LLM structuring yields **1,042 artist** records spanning six roles: composers, choreographers, designers, muses, librettists, and regisseurs. Roles are compiled into a structured dataset and validated against a manually verified subset.

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

NETWORK CONSTRUCTION

The structured data enabled the construction of the artists collaboration network based on the geographic location of their work:

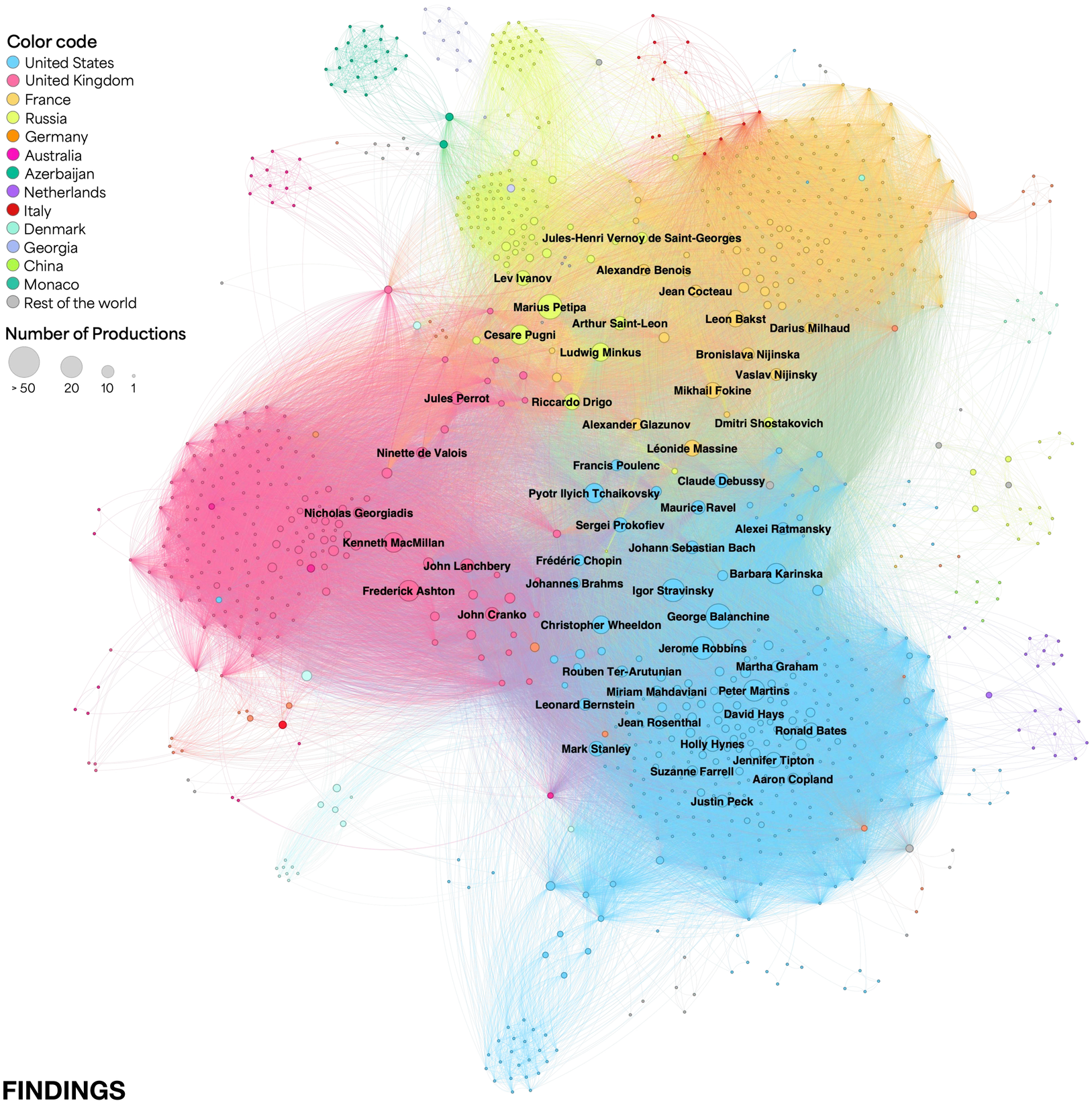


Each artist is represented as a node, and two artists are connected if they share a premiere city. This geographic co-location serves as a proxy for artistic exchange and creative mobility across borders.

NETWORK ANALYSIS

We analyze the structure of the artists network using connectivity and component analysis to assess cohesion across borders, and identify key nodes that bridge creative communities across countries.

AI-structured network of ballet creators



FINDINGS

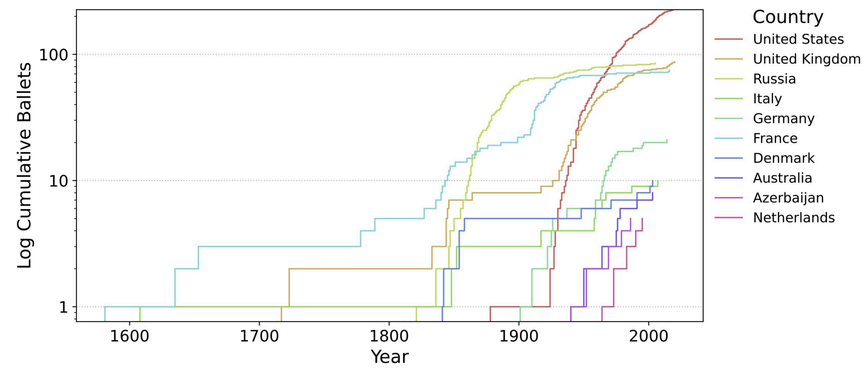
Highly connected network: 99.88% of artists belong to one large connected component.

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.

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. becomes central in ballet production.



Cumulative ballet production over time shows the United States as the dominant center of new creations after a sharp increase between the 1930s and 1940s, coinciding with the emergence of American ballet through the influence of immigrant European artists.

CONCLUSIONS

Using Wikipedia integrated with LLMs, we convert narrative entries into structured records to map creative mobility at scale, linking artists and places in ways that would be difficult to achieve by human annotators alone. This pipeline opens AI-driven pathways for systematic network analysis of human mobility and cultural diaspora, and through more complex multi-agent designs, it can be extended to other art forms and languages.

References

Adamson, F. (2012). *Constructing the diaspora*. In Politics from afar. Columbia University Press.

Schich, M. et al. (2014). *A network framework of cultural history*. Science, 345(6196), 558-562.

Brown. T. et al. (2020). *Language models are few-shot learners*. NeurIPS, 33, 1877-1901.