An Analysis of Worldwide Language Networks and Cultural Clustering

Fozhan Babaeiyan Ghamsari[1⋆], Anh Le[1⋆⋆], Claudia Rawson[1], Lesly Castellanos[1], Salvador Sandoval[1], Richie Prak[1], Oscar Morales Ponce[1]

California State University, Long Beach, Long Beach, CA 90840, USA
fozhan.babaeiyanghamsari01@student.csulb.edu, anh.le02@student.csulb.edu, oscar.moralesponce@csulb.edu

Abstract. This study explores the global structure of multilingual interactions through the lens of network analysis, aiming to determine whether language relationships are more accurately captured by decentralized networks than by models centered on a single dominant language. Using three complementary datasets—OpenSubtitles movie translations, Wikipedia interlanguage links, and spoken-at-home data from the World Values Survey—we construct weighted language networks and apply community detection, centrality metrics, and clustering analysis. Across all datasets, we observe consistent patterns of linguistic clustering that align with historical alliances, colonial legacies, and migration flows. While English consistently emerges as a central hub, regional clusters such as those in Eastern Europe, Latin America, and Southeast Asia demonstrate substantial structural autonomy. These findings suggest that multilingual network models offer a more faithful representation of linguistic and cultural dynamics than monolingual hierarchies, highlighting the value of graph-based approaches in studying global language systems.

1 Introduction

Language is more than a tool for communication-it is a carrier of culture, memory, identity, and power. As globalization accelerates cross-cultural exchange, understanding how languages interact across societies becomes critical for mapping human connection and cultural influence.

Traditional studies of global language dynamics often focus on dominant languages-such as English, Mandarin, or Spanish-as singular vectors of influence. However, this approach risks oversimplifying the complex and multilingual nature of real-world communication. A more holistic view considers languages not in isolation, but as part of dynamic, interconnected networks shaped by history, politics, media, and migration.

In this study, we ask: Can multilingual language networks offer a more accurate and culturally grounded model of global linguistic relationships than models centered on a single universal language? To explore this, we introduce the concept of a *World Language Network (WLN)*-a network model in which each language is represented as a node, and connections (edges) between languages are assigned weights based on the frequency of translation, co-usage, or content-sharing between them. We hypothesize that these networks will naturally form clusters in ways that reflect geopolitical alliances, colonial histories, and regional proximity.

To test this, we analyze three distinct datasets:

- **OpenSubtitles:** subtitle translation frequencies between 94 languages, capturing media-driven language interaction;
- **Wikipedia Interlanguage Links:** hyperlinks between language editions, reflecting digital knowledge connectivity;
- **World Values Survey (WVS):** spoken-at-home language data across 66 countries, providing insight into real-world linguistic presence.

Using graph-based techniques-community detection (Louvain clustering), centrality analysis (PageRank, betweenness, degree), and network visualization-we examine which languages function as hubs or bridges, how languages group into communities, and what cultural patterns emerge from their structure.

⋆ These authors contributed equally. Corresponding author: fozhan.babaeiyanghamsari01@student.csulb.edu
⋆⋆ These authors contributed equally. Corresponding author: anh.le02@student.csulb.edu

Our results suggest that while English consistently plays a central role, multilingual clusters rooted in colonial legacies, migration corridors, and regional identity persist. These findings highlight the value of network models in revealing the cultural geography of language on a global scale.

2 Related Work

Languages are not isolated entities but part of evolving systems shaped by communication, history, and power. Network-based approaches have become essential for modeling the global structure of these linguistic systems. Prior research has consistently shown that language interactions-whether through translation, knowledge-sharing, or everyday use-tend to form clusters shaped by cultural, geographic, and political forces.

Ronen et al. [8] laid the groundwork by constructing a global language network from Wikipedia, book translations, and Twitter. They found that languages cluster by historical relationships such as colonization and regional proximity. This early work established that language networks do not reflect global equality but instead reproduce geopolitical and cultural hierarchies.

Building on this, Johansson and Lindberg [6] analyzed Wikipedia interlanguage links and revealed how editorial practices-especially bot-generated content-distort connectivity in certain language editions. Despite these anomalies, they confirmed that interlanguage networks still reflect meaningful cultural and historical proximity among major languages.

To understand how people bridge languages in practice, Esmaeilialiabadi et al. [3] examined translation demand using Google Translate queries. Their network analysis showed that translation flows are not random but closely aligned with colonial legacies, economic ties, and regional communication needs. Languages like English, French, and Spanish emerged as global pivots.

Other researchers have expanded this perspective using demographic or media-based data. Gurevich et al. [4] introduced a cross-country language connectivity index based on shared official and spoken languages. Their findings revealed clusters of countries aligned along linguistic families and colonial histories. Similarly, Dueñas and Mandel [2] found that even the spread of YouTube music videos follows linguistic and regional pathways, suggesting that language remains a key conduit of cultural diffusion-even in globally accessible media.

This cultural clustering effect is also visible in the flow of misinformation. Quelle et al. [7] found that most false claims stay confined within a single language cluster, and rarely cross into linguistically distant groups. Even when content spreads globally, it often does so along linguistic and historical lines.

These studies frequently rely on clustering methods such as the Louvain algorithm [1] to detect community structures within language networks. The availability of multilingual corpora like OpenSubtitles [9] has made it easier to build large-scale language graphs based on real communication data. Hale [5] further highlighted the multilingual reality of the web, showing that many users access or contribute content across multiple languages, reinforcing the importance of studying interconnected language systems.

While these studies each explore meaningful facets of the global language network, they are typically limited to one platform or dataset-Wikipedia, Google, YouTube, or fact-checking archives. In contrast, our study integrates three complementary datasets-subtitle translations, Wikipedia interlanguage links, and spoken-at-home survey data-to examine whether consistent patterns of linguistic clustering emerge across media, digital, and demographic layers. This triangulated approach offers a more comprehensive view of how language reflects, reinforces, and transcends cultural boundaries.

Note: All figures and visualizations referenced in this section are original works created by the authors.

3 Methodology

3.1 Data Sources

To model multilingual relationships, we draw from three large-scale datasets that capture language interaction across different domains:

1. Movie Subtitles (OpenSubtitles). The OpenSubtitles corpus contains aligned subtitle pairs across 94 languages. We use translation pair counts from the 2024 release, collected via the OPUS API, to build a weighted graph-a network where each connection (edge) between two languages is assigned a numerical value

representing the number of translated sentence pairs between them. This dataset captures cultural exchange through media consumption.

Limitations and Biases: The OpenSubtitles dataset reflects the global film and television industry, resulting in overrepresentation of languages with large international media markets-such as English, Spanish, French, and Portuguese-Brazil-and underrepresentation of many minority or regional languages. Subtitle quality varies, with contributions from professionals, amateurs, and machine translations, which may affect the reliability of translation counts. Metadata inconsistencies, missing or duplicate files, and ambiguous language codes can also impact data accuracy. The dataset is skewed toward recent and popular media, with older or niche productions less represented. Edges are weighted by translation volume, which may reflect industry practices or fan activity rather than organic language interaction. For **visualization clarity only**, our network graphs focus on the most connected languages, potentially understating smaller languages in figures, though not in the underlying analysis. All network construction and quantitative analysis use the full, original alignment count data. Thus, our results reflect translation activity present in OpenSubtitles and should be interpreted within the context of media-driven language interaction, not as a complete representation of global linguistic diversity.

2. Wikipedia Interlanguage Links. We extracted interlanguage link data from the top 20 Wikipedia editions by article count, using language edition SQL dumps. Each edge in this network corresponds to a hyperlink from one language's article to its counterpart in another language, capturing cross-lingual knowledge connectivity. The final network is directed-a network where each connection has a direction, indicating a one-way relationship from one language to another-and weighted by link frequency (the number of times articles in one language edition link to another).

Limitations and Biases: The Wikipedia interlanguage link network is shaped by both editorial practices and automated bot activity, leading to overrepresentation of certain language editions (such as Cebuano and Waray) that have high article counts and interlanguage links due to automated content creation, rather than organic community growth. This can inflate connectivity and centrality for those languages, potentially distorting network structure and cluster detection. The network is constructed from the top 20 Wikipedia editions by article count, which means smaller or less active language editions-and thus many minority or under-documented languages-are excluded. Interlanguage links are not always reciprocal, and some editions link out more than they receive, affecting measures like in-degree, out-degree, and PageRank. Edges are weighted by the number of interlanguage links, reflecting editorial and bot activity rather than real-world linguistic or cultural proximity. For **visualization clarity**, network graphs highlight the most connected languages and strongest links, making complex patterns more accessible in large networks. All network construction and quantitative analysis use the full, original interlanguage link data. Results reflect the editorial and digital knowledge connectivity present in Wikipedia and should be interpreted within the context of platform-driven language interaction, not as a complete representation of global linguistic diversity or real-world language use.

3. Spoken-at-Home Languages (World Values Survey). We used Wave 7 of the World Values Survey (WVS), which includes responses from 66 countries. We extracted question Q272: "What language do you normally speak at home?" This data was used to construct a bipartite network-a network with two types of nodes (countries and languages), where edges only connect nodes of different types-with edge weights corresponding to the proportion of respondents per country.

Limitations and Biases: The WVS dataset covers only 66 countries, so many regions and languages-especially indigenous or minority ones-are missing. Some responses are grouped as "Unlisted Language" or "Other," reducing the visibility of linguistic diversity. Sample sizes vary by country, so smaller languages may not appear. The survey asks for the language "normally" spoken at home, which may not reflect all multilingual households. For **visualization clarity only**, we display in our bipartite graph and heatmap only languages spoken by more than 5% of a country's population. This threshold does not affect our data analysis or findings, but simply makes the figures easier to interpret. All network analysis uses the full, original survey data. Thus, our results reflect the surveyed populations and not the entire world. The full list of surveyed countries and reported languages is available in the official World Values Survey Wave 7 codebook and methodology documentation (see Section "Language spoken at home (Q272)" and country/language code lists)[10].

3.2 Graph Construction and Analysis

Each dataset was transformed into a weighted network:

- **Subtitle Network:** Undirected graph-a network where connections between languages have no direction, indicating a mutual relationship-where edges are weighted by the number of aligned subtitle translations between them.
- **Wikipedia Network:** Directed graph where nodes are language editions and edges are weighted by the number of interlanguage links.
- **Spoken Language Network:** Bipartite graph (country–language) projected into a co-occurrence language–language graph, where two languages are connected if they are both spoken in the same country, and the edge weight represents how frequently this co-usage occurs.

All networks were analyzed using Python's NetworkX and visualized with Matplotlib and Gephi.

3.3 Network Metrics and Community Detection

To analyze language prominence and connectivity, we applied the following metrics:

- **Weighted Degree:** The sum of all edge weights connected to a language node, measuring the total volume of interactions or connections that language has with others.
- **Betweenness Centrality:** A metric that identifies bridge languages-those that frequently lie on the shortest paths between other languages in the network, acting as connectors between different language groups.
- **PageRank:** An algorithm that measures the influence of a language by considering both the number and quality of its connections to other influential languages, similar to how Google ranks web pages.

To identify regional or cultural clusters, we applied the Louvain algorithm for community detection-an unsupervised method that groups languages into clusters (communities) so that languages within the same group are more densely connected to each other than to those outside the group. The algorithm seeks to maximize a value called modularity, which quantifies the strength of the division into communities.

4 Results and Analysis

4.1 Subtitle Translation Network (OpenSubtitles)

The subtitle translation network, constructed from the OpenSubtitles corpus, captures global media exchange through multilingual subtitling. In this network, nodes represent languages, and weighted edges denote the number of subtitle-aligned sentence pairs between them.

Figure 1 illustrates a strongly right-skewed (heavy-tailed) distribution: most translation pairs have low volumes, but a few pairs have extremely high volumes, creating a long tail of high values. Most translation pairs have low volumes, while a small number exchange extremely high volumes. The median line count per language pair is approximately 140,000, but the mean exceeds 6 million due to a handful of very strong links. Only a small subset of edges surpasses 10 million lines, forming the network's long tail. This distribution is visualized in Figure 1.

As shown in Figure 2, English functions as the central hub, connected to nearly all major languages. The most translated language pairs are:

- English →Portuguese (Brazil) ∼115M lines
- English →Spanish ∼105M lines
- English →Romanian ∼100M lines
- English →Arabic ∼88M lines
- English →French ∼84M lines

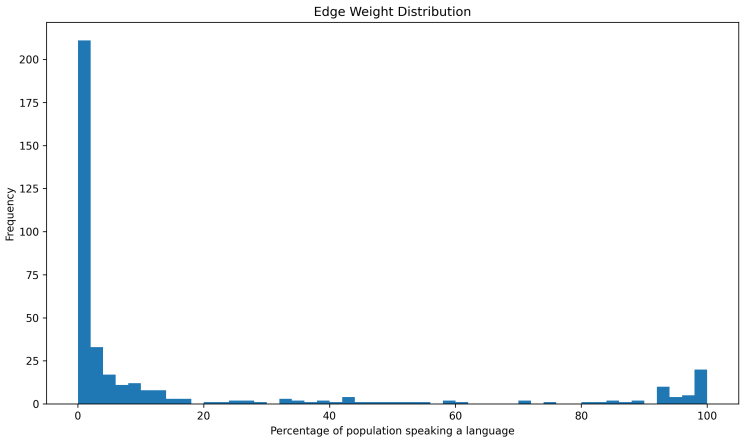

Fig. 1. Edge weight distribution (log scale). A few translation pairs dominate the network, showing a heavy-tailed distribution.

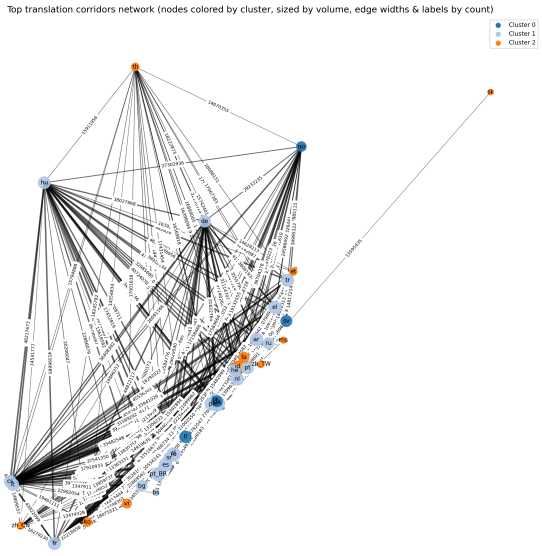

Fig. 2. Top translation corridors in the subtitle network. Node size represents weighted degree (languages with more or stronger connections appear larger), and node color indicates the Louvain cluster (community) to which each language belongs. Edges are weighted by subtitle alignment volume.

Figure 3 summarizes key centrality metrics. English scores highest in all three-PageRank, betweenness, and weighted degree-reinforcing its role as a bridge language, connecting different parts of the network and facilitating communication between otherwise separate language groups. Portuguese (Brazil) and Romanian also exhibit surprisingly high centrality, likely due to regional subtitling practices and the influence of diaspora communities.

The clustering structure is further illustrated in Figures 4.1–4.1. Louvain clustering reveals three major language clusters:

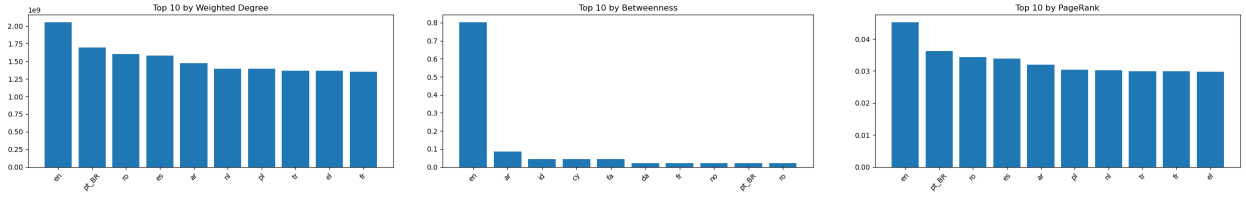

Fig. 3. Top 10 languages ranked by centrality metrics. Left: Weighted Degree. Center: Betweenness Centrality. Right: PageRank.

- **Cluster 0 (dark blue):** A global cluster centered on English and other high-traffic languages. *Historical ties: colonialism; economic ties: global trade; cultural ties: media distribution (e.g., Hollywood).*
- **Cluster 1 (light blue):** Western and Slavic European languages with strong mutual connections. *Political ties: EU membership; economic ties: European commerce.*
- **Cluster 2 (orange):** Scandinavian languages with dense internal connections. *Political legacy: Union of Kalmar; modern institutions: Nordic Council.*

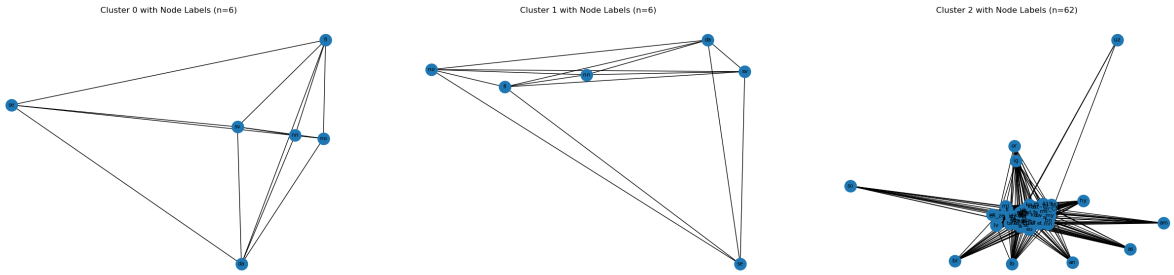

Fig. 4. Left: High-traffic global languages. English acts as the core of a multilingual, media-driven network. Center: Western and Slavic European languages. Characterized by tight mutual translation activity. Right: Scandinavian languages. Internal coherence reflects shared cultural and political history.

These results confirm that multilingual networks are structured by cultural, political, and economic histories. Despite English's global dominance, regionally coherent clusters persist-suggesting that network-based language modeling can reveal both global centrality and localized cohesion.

4.2 Wikipedia Interlanguage Link Network

The Wikipedia interlanguage network models how knowledge is shared across language editions. In this directed, weighted graph, nodes represent Wikipedia language editions, and an edge from language A to B indicates the number of articles in A that link to their equivalents in B.

Figure 5 shows that English is the largest and most linked language edition, receiving links from nearly all other editions. This reflects its role as a central source of knowledge on the platform. French, Spanish, and German also appear as strong hub languages with many connections, acting as central points in the network-with high mutual link exchange.

The resulting language co-occurrence network is visualized in Figure 9. Louvain clustering reveals several distinct communities:

- A **core global cluster** (green) includes English, French, Spanish, German, and Russian-well-developed editions with many mutual links.
- A **peripheral cluster** (purple) includes languages like Cebuano, Waray, and Dutch. These editions are inflated by automated article creation (e.g., bots), which increases article count but not mutual connectivity.

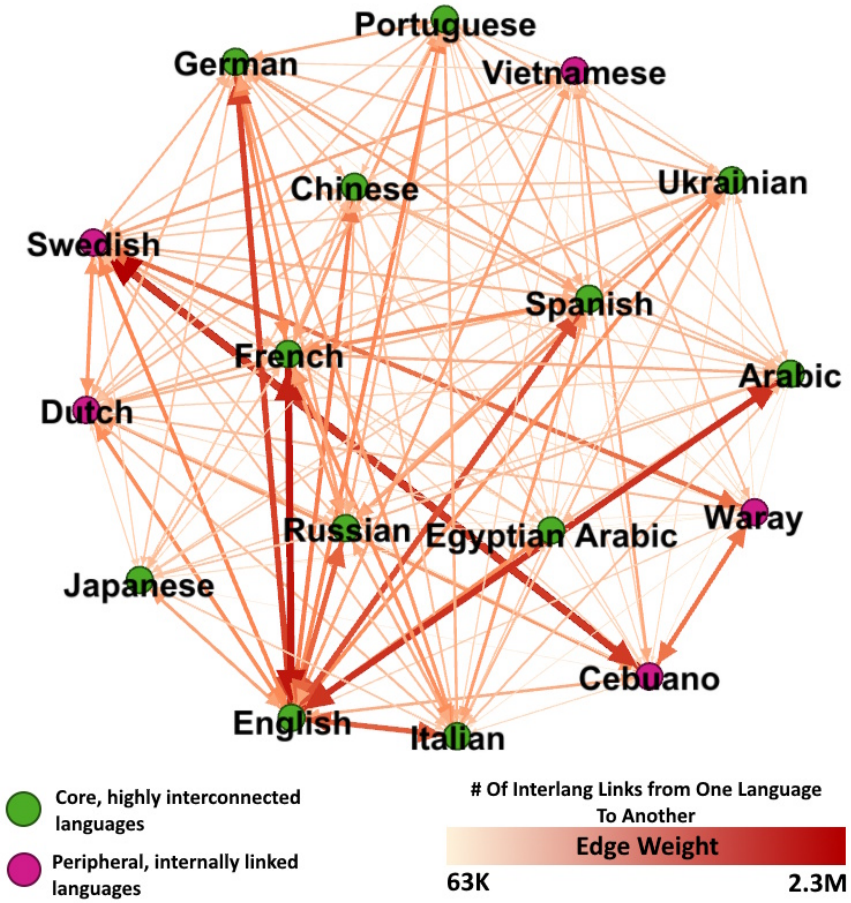

Fig. 5. Wikipedia interlanguage network. Node size corresponds to in-degree; edge thickness indicates link volume. Node color shows Louvain cluster.

In Figure 6, when the same network is geographically overlaid, regional proximities align with higher interlanguage connectivity. For instance, strong linkages are visible between Spanish, Portuguese, and Catalan; and between Germanic languages such as German, Dutch, and Swedish.

These patterns reflect not only linguistic similarity but also political and historical relationships. For example, English and French both act as hubs for many African and Southeast Asian language editions due to colonial history and institutional partnerships.

Compared to the subtitle network, the Wikipedia network reflects more editorial and infrastructural asymmetries. Bot-driven editions distort standard network measures like degree and clustering. Nevertheless, community detection still reveals cultural blocks shaped by both language family and content strategy.

This network reinforces the observation that global language systems are not strictly hierarchical. While English remains central, the presence of mutually reinforcing regional hubs illustrates the decentralized structure of digital knowledge production.

4.3 Spoken Language Network (World Values Survey)

The spoken language network constructed from the World Values Survey (WVS) captures ground-level linguistic distributions across 66 countries. Each respondent answered question Q272, "What language do you normally speak at home?", offering a direct insight into household language use, independent of media exposure or institutional status.

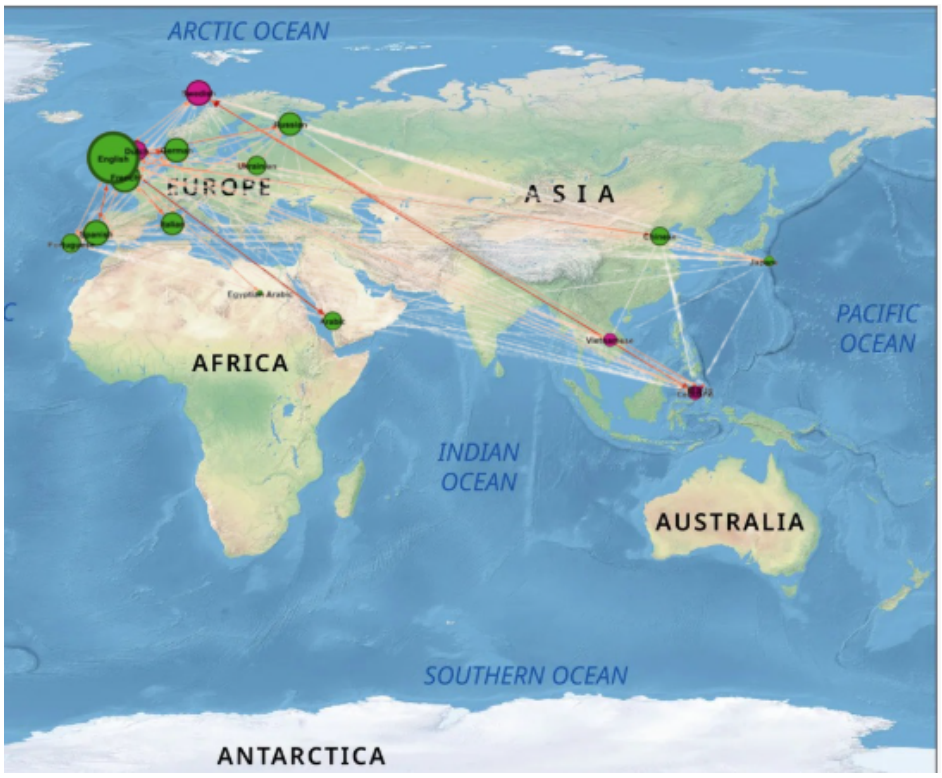

Fig. 6. Wikipedia interlanguage network mapped geographically. Stronger links appear between culturally or regionally aligned languages.

A bipartite network was built connecting countries to languages, with edge weights corresponding to the percentage of survey respondents who reported speaking a particular language. For readability, Figure 7 filters out edges where fewer than 5% of a country's population speaks a language.

As shown in Figure 7, many countries link to a single language, such as Bengali in Bangladesh or Arabic in Egypt, indicating national linguistic homogeneity. In contrast, countries such as India, Nigeria, and the Philippines link to multiple languages, revealing rich internal linguistic diversity and multilingualism shaped by colonial history, ethnicity, and migration.

To further analyze language prominence, we visualized the edge weight distribution (Figure 8). The histogram shows how frequently different language percentages appear across countries. The distribution is heavily right-skewed: while a few country–language pairs reach 100% usage, the majority fall below 10%, representing minority or regional languages.

We then projected the bipartite network into a language-language graph. Two languages are connected if they co-occur in the same country, with edge weights proportional to shared speaker populations. Louvain community detection was applied to reveal network communities (Figure 9).

The Louvain algorithm revealed distinct language clusters:

- **Global / Colonial Cluster (purple)**: Includes English, Arabic, French, Urdu, and Hindi-languages spread through colonization, migration, and trade.
- **Southeast Asian Cluster (blue)**: Javanese, Cebuano, Tagalog, and other languages from Indonesia and the Philippines.
- **Eastern European / African Cluster (green)**: Russian, Slovak, Ukrainian, Swahili, Luo-languages tied through regional or political alignment.
- **Tribal / Southern African Cluster (yellow)**: Tshivenda, Tonga, Shona, Chitoko-indigenous African languages with strong national ties.
- **Isolated / Other (gray)**: Peripheral or low-co-occurrence languages, such as Romansh or Assyrian Neo-Aramaic.

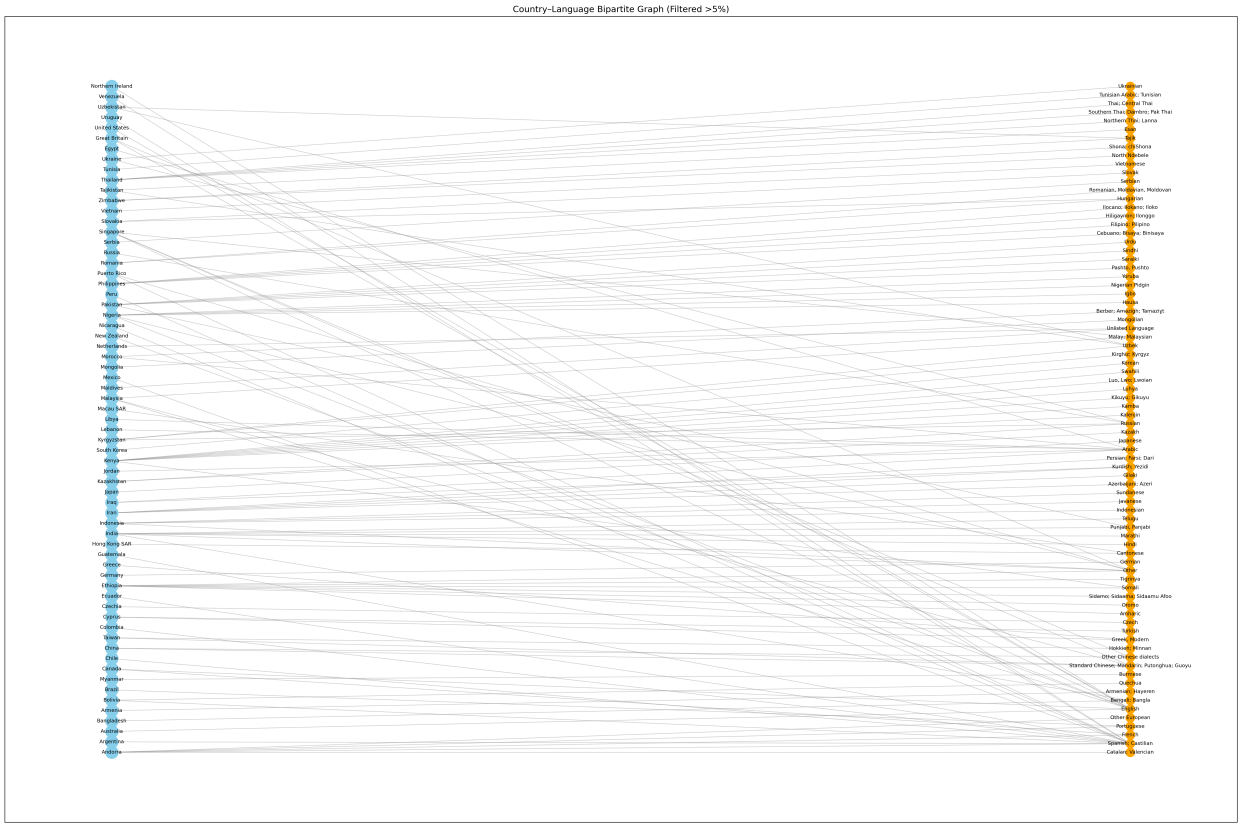

Fig. 7. Country–Language Bipartite Graph (filtered for usage $> 5\%$). Countries are shown in blue, languages in orange.

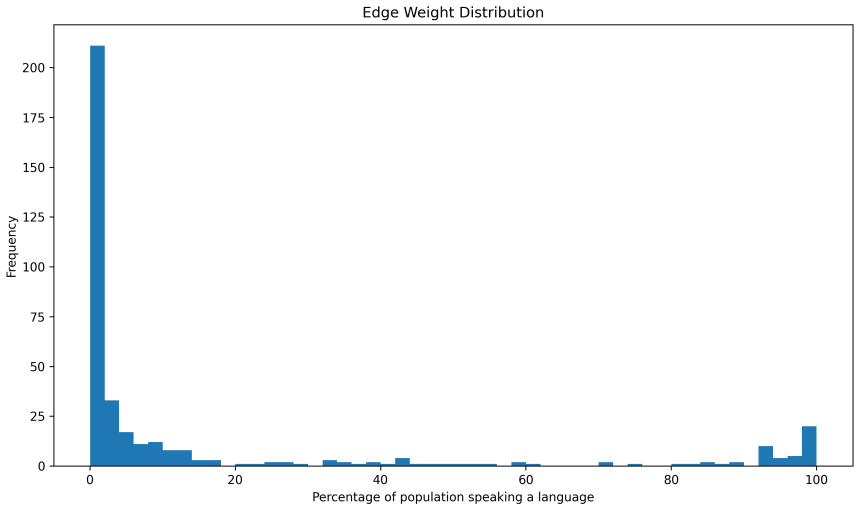

Fig. 8. Edge Weight Distribution: Percent of population speaking a language in each country. Most connections represent small language populations.

At the center of the network lies a dense cluster of globally spoken languages, with English as the most prominent hub. It connects directly to Spanish, Arabic, Hindi, and Mandarin, underscoring its bridging role

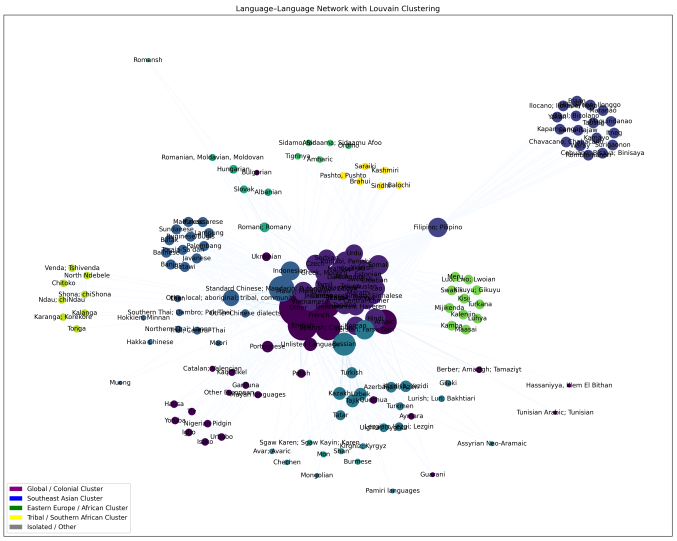

Fig. 9. Language–Language Network with Louvain Clustering. Node size indicates weighted degree; node color shows cluster membership.

across diverse language groups. Surrounding this core are several regional clusters: a Romance group (French, Spanish, Portuguese, Italian), a Slavic cluster (Russian, Ukrainian), and a Southeast Asian cluster (Tagalog, Cebuano, Filipino).

Central Asian languages such as Kazakh, Uzbek, and Kurdish also group together, reflecting geographic and cultural proximity. Interestingly, Bulgarian appears close to Slovak and Hungarian geographically but is structurally clustered with Russian and Romanian due to stronger co-use patterns. Arabic appears within the European cluster-likely influenced by modern migration trends that have increased Arabic usage in Europe.

Smaller, less globally dominant languages are still represented. Some link to regional hubs, while others connect directly to English, highlighting the diverse and interconnected nature of home language usage.

To visualize the intensity of language use, we constructed a heatmap of the top 30 global languages across all countries (Figure 10).

This heatmap reveals clear national dominances: Spanish across Latin America, Arabic in North Africa and the Middle East, and English's scattered presence across regions. Countries like India and South Africa display multiple active languages, confirming multilingual realities observed in the bipartite structure.

Overall, the spoken language network reveals both expected and surprising connections between languages, shaped by population distribution, history, and migration. The structure complements findings from other networks and underscores the value of network analysis in modeling real-world linguistic diversity.

4.4 Cross-Network Synthesis and Cultural Patterns

Across all three networks-subtitles, Wikipedia links, and spoken-at-home data-we observe a consistent tendency for languages to cluster according to historical, cultural, and geopolitical factors. These clusters are not artifacts of any one medium or dataset, but appear repeatedly in different linguistic contexts: mass media, digital encyclopedias, and everyday communication.

English consistently emerges as the most central node across all networks, acting as a global linguistic bridge. However, its dominance is not uniform. In the spoken-at-home network, for example, English often shares space with regional languages, suggesting that cultural presence does not always translate into grassroots usage.

Regional clusters also reappear across all layers:

– A *Scandinavian cluster* (e.g., Swedish, Danish, Norwegian) is tightly connected in subtitles and Wikipedia, reflecting historical and political cooperation.

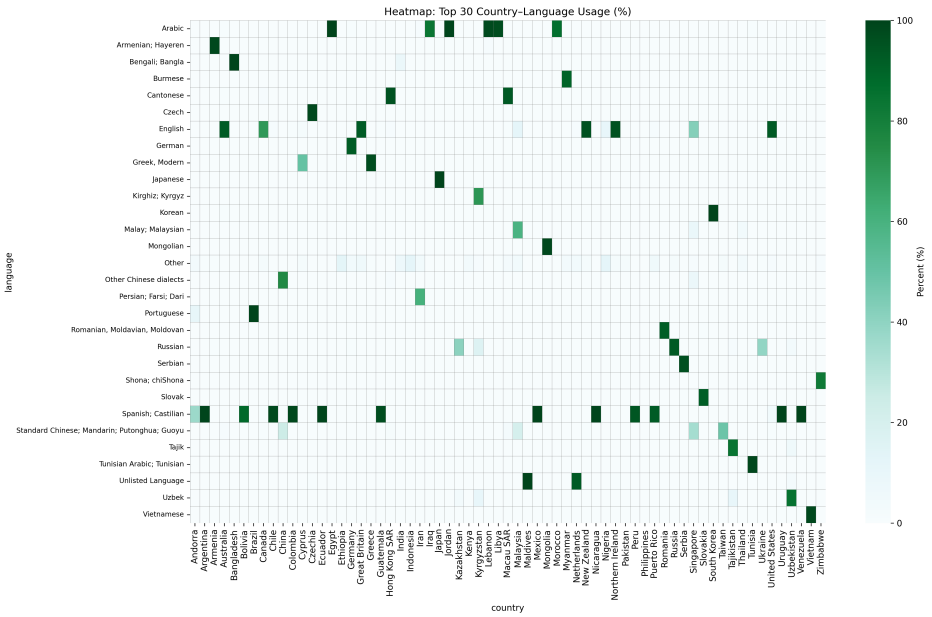

Fig. 10. Heatmap: Top 30 Spoken Languages Across Countries (WVS Q272). Darker shades represent higher usage percentages.

- *Western European and Slavic languages* form overlapping communities, often shaped by EU membership, cross-border media, and educational exchange.
- A *global colonial cluster* includes English, French, Arabic, and Portuguese, often reflecting former empire languages used for administration, trade, or migration.

Interestingly, some languages exhibit strong centrality in one network but not others. For instance, Romanian has high translation volume in the subtitle network but plays a less prominent role in Wikipedia or WVS data. This highlights how platform-specific dynamics, such as media subtitling norms, can shape overall structure and patterns of connections in the network.

Overall, the convergence of structure across these independent networks suggests that language relationships are deeply embedded in global cultural systems. A network-based model not only captures central hubs but also reveals enduring regional cohesion. Our findings underscore the value of triangulating diverse datasets to reveal the multi-layered geography of global language interaction.

5 Conclusion and Future Work

This study introduced a multi-layered approach to analyzing global language relationships through the lens of network science. By constructing and comparing language networks derived from subtitle translations, Wikipedia interlanguage links, and spoken-at-home data, we identified consistent patterns of cultural clustering, regional cohesion, and linguistic centrality.

Across all datasets, English consistently emerged as a global bridge language, while other colonial or regionally dominant languages-such as Spanish, Arabic, French, and Portuguese-formed secondary hubs. Importantly, community detection revealed recurring clusters that reflect historical, political, and cultural alignments, such as Scandinavian cooperation, post-colonial linguistic blocs, and multilingual cross-border regions in Europe and Asia.

Our results demonstrate that a single “universal language” cannot fully capture the diversity of global linguistic systems. Instead, multilingual network models better reflect how languages function in parallel-shaping, preserving, and transmitting culture across digital, institutional, and everyday contexts.

Looking ahead, several directions for future research are worth exploring. One avenue is to incorporate additional platforms such as social media (e.g., Twitter, Reddit, YouTube) or multilingual news corpora to

observe how real-time language interaction differs from formal content like Wikipedia or survey responses. Another extension could focus on temporal dynamics-measuring how language clusters evolve over time in response to geopolitical change, migration, or digital media trends.

Finally, this type of cross-network linguistic modeling could inform the development of more equitable and culturally aware multilingual AI systems, including translation tools, voice interfaces, and search algorithms. As global communication continues to grow, so does the need for computational methods that reflect not just linguistic frequency, but linguistic context, history, and cultural proximity.

References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10), P10008 (2008)
2. Dueñas, M., Mandel, A.: The structure of global cultural networks: Evidence from the diffusion of music videos. PLOS ONE **18**(11), e0294149 (2023)
3. Esmaeilialiabadi, D., Avşar, B., Yousefnezhad, R., Aliabadi, E.E.: Investigating global language networks using google search queries. Expert Systems with Applications **121**, 66–77 (2019)
4. Gurevich, T., Herman, P.R., Toubal, F., et al.: A dataset on linguistic connectivity across and within countries. Scientific Data **12**, 542 (2025)
5. Hale, S.A.: Global connectivity and multilinguals on the web. Information, Communication & Society **17**(4), 405–421 (2014)
6. Johansson, S., Lindberg, Y.: Wikipedia as a virtual learning site and a multilingual language site. In: Ahlqvist, S., Olsson, M. (eds.) Virtual Sites as Learning Spaces, pp. 181–204. Springer, Cham (2019), `https://www.diva-portal.org/smash/get/diva2:1426913/FULLTEXT01.pdf`
7. Quelle, D., Cheng, C.Y., Bovet, A., et al.: Lost in translation: using global fact-checks to measure multilingual misinformation prevalence, spread, and evolution. EPJ Data Science **14**(22) (2025)
8. Ronen, S., Gonçalves, B., Hu, K.Z., Vespignani, A., Pinker, S., Hidalgo, C.A.: Links that speak: The global language network and its association with global fame. In: Proceedings of the National Academy of Sciences. vol. 111, pp. E5616–E5622. National Academy of Sciences (2014)
9. Tiedemann, J.: Parallel data, tools and interfaces in opus. Proceedings of LREC pp. 2214–2218 (2012)
10. World Values Survey Association: World Values Survey Wave 7 (2017–2022): Codebook and Methodology (2022), available at: `https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp`