Mapping the Armenian Diaspora Network

Garo Kerdelian[1,2]

1. Network Science Institute, Northeastern University, Boston, MA 02115
2. Department of Physics, Northeastern University, Boston, MA 02115

The Armenian diaspora is one of the largest and oldest diasporas in the world [1]. While Armenians have migrated for centuries to trade, find labor, or escape conflict, the biggest wave of Armenian migration occurred forcibly during the Armenian Genocide, the systematic deportation and slaughter of around 1.5 million Armenians by the Ottoman Empire [2]. Starting in 1915, the Ottoman Turks began forcibly displacing and killing hundreds of thousands of ethnic Armenians, reducing the number of Armenians in what would become Turkey by 90 percent [3]. Survivors of the Genocide arrived in Syria and Lebanon; some settled there for generations and others found new homes around the world [4]. Over a century from the Genocide, global Armenian communities preserve the homeland's cultural practices, language, and identity, maintain a rhetoric of collective trauma and mourning, and sustain contact with each other and modern Armenia, establishing diasporas rather than dispersed ethnic communities [5].

The global Armenian diaspora is estimated at around 7 million people today distributed in over 100 countries. Figure 1a shows notable diaspora centers: Russia is estimated to host the largest Armenian diaspora population, roughly matching that of Armenia itself (2.7–3 million), and the United States the second largest, estimated at 1.2–1.6 million [6]. Beyond population estimates, however, the Armenian diaspora remains largely unstudied as a network. The central challenge to mapping diaspora connections is data availability. On one hand, many countries adopt "colorblind" policies that prohibit ethnic and racial classification, so surveys cannot be used to enumerate Armenian diaspora populations. On the other hand, diaspora members practice identity in diverse, context-specific ways: some speak Armenian as a first language while others speak the language of the host country; some preserve identity through cultural practices while others do so through religious ones; many assimilate into the host country altogether after one or two generations. Therefore, the evolution of the Armenian diaspora network is a unique historical phenomenon, yet mapping it is a massive undertaking.

In this study, we propose a starting point to mapping the Armenian diaspora network through Wikipedia hyperlinks and aim to answer the following questions: how are diasporas connected within and among each other? Is the network more structured than expected by chance? Is the global diaspora divided into insular groups? Uncovering diaspora connections is essential to understanding its structure, interconnectedness, and impact—a question for network science.

The nodes of our network are Armenian individuals with Wikipedia pages collected from three sources: (1) a Wikidata query returning individuals with a categorical connection to Armenian identity, such as Armenian citizenship, Armenian Apostolic/Catholic faith, Armenian ethnicity, and many others (2) manual scraping of Wikipedia "Category" pages such as *Lebanese people of Armenian descent* and *American people of Armenian descent*, and (3) a Claude query returning a list of Armenian individuals with Wikipedia pages. The three sources are harmonized into 4,946 unique entries, each with a Wikidata ID, name, Wikidata URL, and Wikipedia URL; Wikidata IDs serve as node IDs. An undirected, unweighted link is created between two nodes if the Wikipedia page of one has a hyperlink to that of the other, which suggests a documented relationship between the two individuals.

We analyze the resulting network of 4,946 nodes and 11,189 links using tools from network science. With an average degree of $\langle k \rangle = 4.5$ and a second moment of $\langle k^2 \rangle = 142.0$, the network is highly heterogeneous in its connectivity. The log-binned degree distribution, the probability that a randomly chosen node has k connections, has an approximately straight-line form on log-log axes (Figure 1b), suggesting consistency with a power law, $p_k \propto k^{-\gamma}$, as found in many social, biological, and technological networks [7]. A fit yields $\gamma \approx 2.26$, within the typical $2 < \gamma < 3$ regime, consistent with a scale-free organization: most Armenian individuals in the network have few connections, while a few hubs are very well-connected across the diaspora.

We next examine clustering, shortest paths, and degree assortativity, comparing each observed value to the mean over 10 degree-preserving randomizations of the network [8]. The network shows higher clustering, longer shortest paths, and stronger (positive) assortativity than expected at random, with z-scores of 117, 132, and 41 respectively (Figure 1c). High clustering and positive assortativity—high-degree nodes preferentially linking to other high-degree nodes—are characteristic of social networks, while paths longer than the randomized baseline suggest strong modular organization.

To verify the hypothesis, we run the Leiden community-detection algorithm [9] on the largest connected component of the network and identify 28 communities. The network has a modularity of 0.775, far above the average of 0.299 obtained for 10 degree-preserving randomizations. The community-size distribution is heavy-tailed, spanning communities from 3 to 528 members. Manually inspecting the communities reveals coherent groupings: politicians, musicians, soccer players, and post-Soviet figures. Figure 1d visualizes the largest connected component with community label annotations.

Our preliminary results suggest a diaspora network with scale-free features, high clustering, and well-separated communities. Our analysis is naturally skewed towards public figures rather than the broader diaspora given our use of Wikipedia

data. We hope to extend the graph with more data sources and relationship types, validate community assignments against diaspora organizations, and study the growth of the Armenian diaspora as a temporal network.

[1] International Organization for Migration. (2021). Skills mapping through big data: A case study of Armenian diaspora in the United States of America and France. International Organization for Migration.
[2] Kévorkian, R. (2011). The Armenian Genocide.
[3] Suny, R. G. (2015). "They Can Live in the Desert but Nowhere Else": A History of the Armenian Genocide: A History of the Armenian Genocide. Princeton University Press.
[4] Migliorino, N. (2022). (Re) constructing Armenia in Lebanon and Syria: Ethno-cultural Diversity and the State in the Aftermath of a Refugee Crisis. Berghahn Books.
[5] Tölölyan, K. (2007). The contemporary discourse of diaspora studies. Comparative Studies of South Asia, Africa and the Middle East, 27(3), 647-655.
[6] Office of the High Commissioner for Diaspora Affairs. Armenian diaspora communities. https://diaspora.gov.am/en/diasporas
[7] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. science, 286(5439), 509-512.
[8] Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. Science, 296(5569), 910-913.
[9] Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 9(1), 5233.

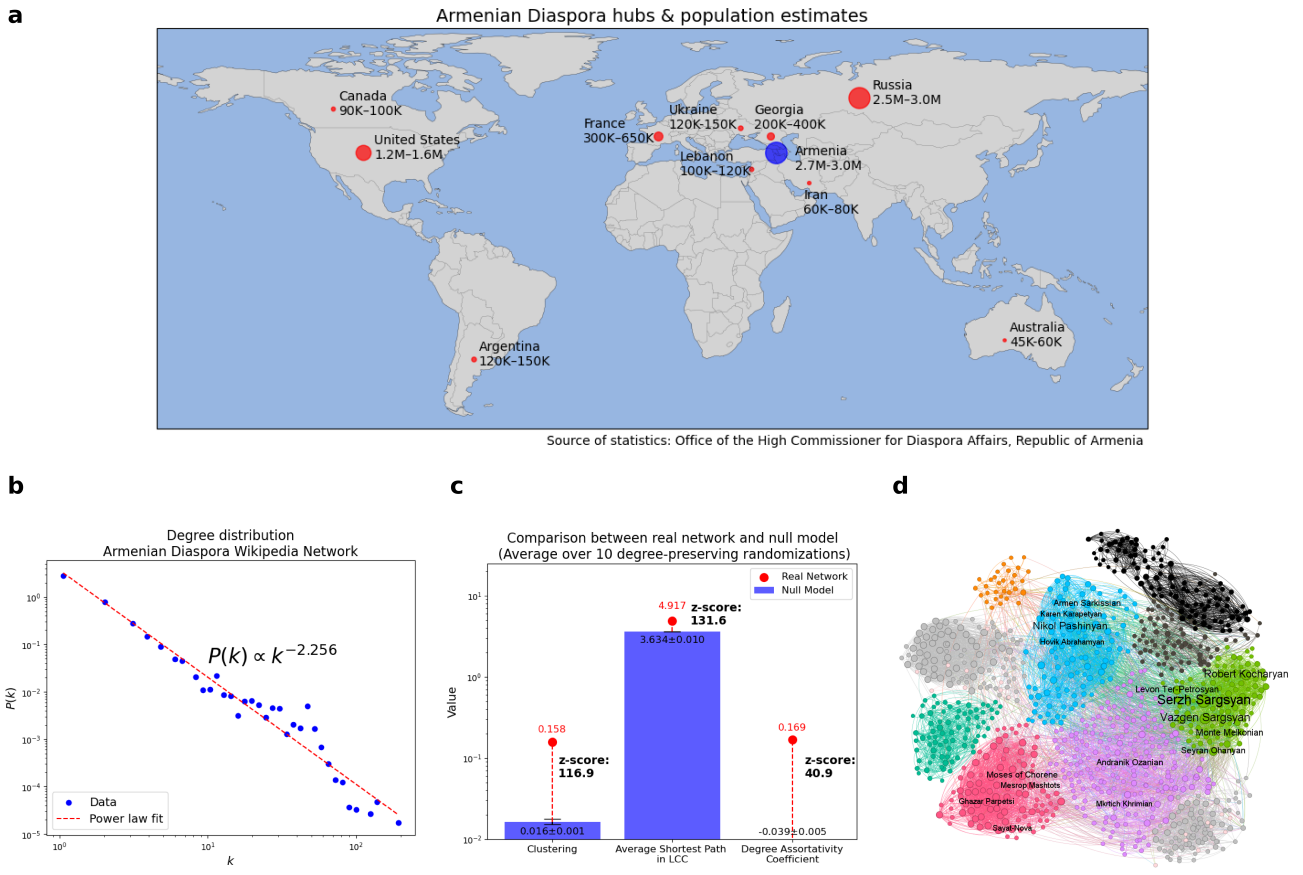


Figure 1: **Mapping the Armenian Diaspora Network through Wikipedia hyperlinks. (a)** The Armenian diaspora comprises around 7 million people today in over 100 countries, with population centers in Russia, the United States, France, and Georgia. **(b)** The degree distribution of the Armenian Diaspora Wikipedia Network is heavy-tailed. A simple power-law fit of the data yields an exponent of 2.26 in the typical exponent range between 2 and 3 of scale-free networks. **(c)** The network has higher clustering, longer shortest paths, and higher assortativity in comparison to averaged degree-preserving randomizations (last bar omitted - null-model value is negative, approximately zero as expected). **(d)** All nodes in the largest connected component of the network with degree larger than 4 are visualized. Node size is proportional to degree and node color corresponds to community label. High-degree node labels are shown.