Using Wikipedia data to monitor big migration events: opportunities and challenges

Carolina Coimbra Vieira*, Ebru Sanlitürk, Emilio Zagheni
[1]Max Planck Institute for Demographic Research
coimbravieira@demogr.mpg.de

Introduction

Migration flows are challenging in terms of tracking and predicting. The challenge increases when these migration events happen due to unforeseen factors like conflicts, resulting into forced migration (Cesare et al. 2018; Leurs and Smets 2018; Tjaden 2021). Digital trace data proposes an alternative source to predict and adapt to real-world changes faster than traditional data sources. Digital trace data from Google Trends (Böhme, Gröger, and Stöhr 2020), Facebook Ads (Spyratos et al. 2018; Alexander, Polimis, and Zagheni 2019) and LinkedIn Ads (Zhu, Fritzler, and Orlowski 2018; Perrotta et al. 2022; Vieira et al. 2022) have been shown as good data sources to track and predict migration. However, in this work, we explore data from Wikipedia.

Wikipedia is the largest and most popular free online encyclopedia, aiming to provide equal access to information about current events and media coverage of a topic worldwide (Singer et al. 2017). Wikipedia pages are created and edited by volunteers around the world (Graham and Dittus 2022) and volunteer contributions to Wikipedia can be affected by events in the real world. For instance, the death of Queen Elizabeth II, on 8 September 2022, and the rapid reaction of Wikipedia editors to update Wikipedia pages related to the queen and the Royal family[1] is a good example of how zealous Wikipedia editors are making quick changes in response to real-world changes.

In this study, we aim to shed light on the relationship between online sources of information and migration flows as well as migrant networks. Forced migration, due to its nature, increases the need to access information quickly and efficiently, and online sources are known to help refugees to meet that need (The GSM Association 2017; Ulutürk, Uysal, and Varol 2019). Wikipedia is also a worldwide well-known source of information. By examining the association between information changes on Wikipedia by language and recent refugee flows, we aim to determine (i) whether Wikipedia data can be used as a predictor for mass migration flows and (ii) how the accumulation of information on Wikipedia relates to migrant networks in a country.

[1]https://www.npr.org/2022/09/15/1122943829/wikipedia--queen-elizabeth-ii-death-deaditors-editors-article

The contribution of this study is considered twofold. First, we introduce Wikipedia data as a novel source to analyze mass migration flows, exploiting the increased need for information and condensed information-seeking patterns. Second, we determine the timing of information changes on Wikipedia and mass migration events, thus showing how it can be used as a predictor for ongoing mass migration flows as well as a broader proxy for information sharing in migrant networks.

Our methodology consists of using Wikipedia data to assess how the number of views on Wikipedia pages dedicated to cities around the world change across time in response to migration events. As a case study, we focus on a recent migration event – the Ukrainian refugee crisis in 2022 – caused by the Russian invasion of Ukraine. For each Wikipedia page dedicated to the capitals and most populous European cities, we investigate the number of views in Ukrainian (i.e., the official language in Ukraine), Russian, and English.

Our results show an increase in the number of views on Ukrainian Wikipedia pages dedicated to Polish cities in 2022. When we contrast these numbers with official data provided by the United Nations High Commissioner for Refugees (UNHCR), we observe a strong association between the number of Ukrainian refugees crossing the border from Ukraine to Poland and the number of views on Ukrainian Wikipedia dedicated to Polish cities. We contribute to the literature on the use of digital trace data to study forced migration in real-time. Our results reveal opportunities in the use of Wikipedia as a proxy to study and predict mass migration flows.

Related Work

Monitoring migration events

Online search engines provided a useful tool to measure interest in a topic that can be used as a proxy for intention to move in migration studies. Availability of online search data, thus, paved the way for a growing literature to predict and analyze migration processes (Lin, Cranshaw, and Counts 2019; Böhme, Gröger, and Stöhr 2020; Avramescu and Wiśniowski 2021; Golenvaux et al. 2020). However, there are some limitations involving the use of online search engine data. One limitation is the availability of various search engines and the difficulty of comparing online search

data from different sources. While Google is the most popular search engine worldwide, Bing and Yandex also appear as competitors in different regions. Each search engine reports online search interest in its own way, using different parameters and algorithms. These algorithms may create a bias themselves, for instance, Google Trends provides only a normalized index for the given place and time and applies an unobservable threshold that prevents results for very low interest. Lack of access to logs containing absolute numbers of searches by Google Trends constitutes a significant shortcoming for detailed comparative research using the most popular search engine in the world. In contrast, Wikipedia data regarding the number of views on some specific pages are easily accessible through the API. Previous work has shown the high correlation between frequently searched keywords and Wikipedia page views, suggesting that Wikipedia page views can a source to determining popular global web search trends (Yoshida et al. 2015). Additionally, Wikipedia is an interesting complementary data source, especially when traditional surveillance systems are not available in real time (Vilain et al. 2017).

Studies using Wikipedia

Wikipedia is the most popular free online encyclopedia maintained by volunteers experts in some domains (West, Weber, and Castillo 2012; Agarwal et al. 2020) around the world (Graham and Dittus 2022; Panciera, Halfaker, and Terveen 2009). Wikipedia readers can benefit from free access to information about current events and media coverage of a topic (Singer et al. 2017). For instance, a large percentage of Wikipedia users read about entertainment-related topics (e.g., TV series, movies, and biographies), history, health, and tech content (Lehmann et al. 2014).

English articles are the most notably considering the number of collaborative interactions to create and edit them. However, Wikipedia is available in 288 other active languages (Bipat, McDonald, and Zachry 2018). Although many articles are available in many languages, some concepts are not represented or not shared across languages (Miquel-Ribé and Laniado 2020, 2019). Other topics, such as some historical figures, appear in multiple languages depending on interactions between cultures (Eom et al. 2015). The number of edits and the complexity also vary according to the language. Overall, editors are less likely to edit complex topics in a second language, except in English since the level of complexity of English edits is the same regardless of the primary language of the editor (Kim et al. 2016). Based on the fact that some specific languages will be edited more often by editors who spoke the language as a primary language, our study uses the number of edits in one language as a proxy for the number of users from a region where people speak the language living in the city or country mentioned on the Wikipedia page.

The number of contributions on Wikipedia can also be affected by events in the real world. For instance, during COVID-19 mobility restrictions, the number of contributions to Wikipedia increased (Ruprechter et al. 2021), especially in languages associated with countries where the most severe mobility restrictions were implemented (Ribeiro et al.). Moving to another topic, Twyman, Keegan, and Shaw (2017) also showed how movements such as *Black Lives Matter* are quickly reported and updated on Wikipedia. As a result of this process, Wikipedia works as a repository of knowledge about social movements as they unfold. In politics, Agarwal et al. (2020) showed that both edits and readership across Wikipedia pages dedicated to politicians in the UK are affected during election times. In finances, Moat et al. (2013) showed that the number of views on Wikipedia pages sign moves on the stock market. Finally, in public health, Vilain et al. (2017) used Wikipedia data to monitor the trends of seasonal diseases in France. In this sense, we believe that Wikipedia can be considered an interesting and important complementary data source, especially when traditional surveillance systems are not available in real time.

Finally, related to migration, Lucchini, Tonelli, and Lepri (2019) collected data on Wikipedia regarding historically notable individuals' movements in order to study important features of migration's behavior. However, the work is limited to content analysis and a specific set of notable individuals. Lerner and Lomi (2019), on the other hand, shows the network effects on the number of edits on migration-related topics on Wikipedia. Although the topic is migration, the study focuses on network effects on edits, while in our study we want to investigate the usefulness of Wikipedia to predict migration flows and as a proxy for information sharing in migrant networks. To the best of our knowledge, this is the first study to use Wikipedia page views to monitor migration flows.

Data and Methods

Wikipedia page views

We collected the data regarding the number of views on Wikipedia pages using the Wikimedia pageviews API[2]. The API provides daily counts for the number of views (i.e., every time the page is loaded) on each Wikipedia page across different languages since July 2015. We only consider user views, including editors, anonymous editors, and readers. Views from search engine "web crawlers" or automated programs are not included.

Figure 1 shows the total number of views across languages on Wikipedia pages dedicated to the 334 European capitals[3] and European cities with more than 200,000 inhabitants within the timeframe of the Eurostat database[4]. The European Wikipedias with more views are in English, German, Spanish, French, and Russian Wikipedia. The Ukrainian Wikipedia is in 22nd place on this list of Wikipedia pages sorted by the number of views.

UNHCR data

We collected data from the United Nations High Commissioner for Refugees (UNHCR) to contrast it with the data collected on Wikipedia. We use this data to assess the level of association between the number of Ukrainian refugees

[2]https://pageviews.wmcloud.org/langviews/

[3]https://en.wikipedia.org/wiki/Category:Capitals_in_Europe

[4]https://ec.europa.eu/eurostat/data/database

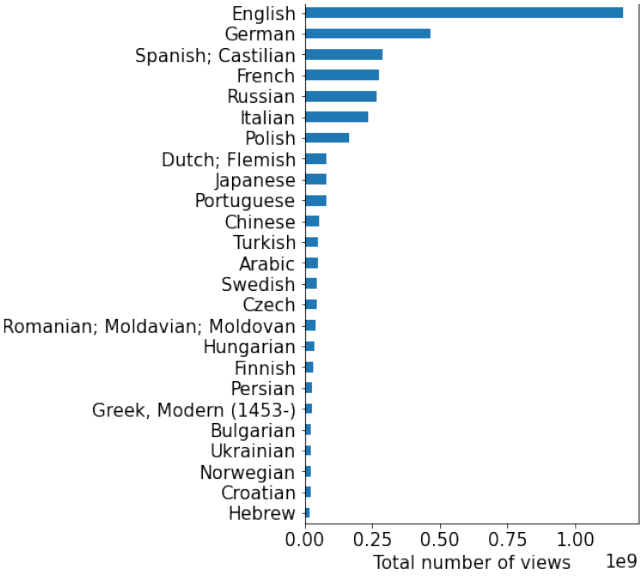


Figure 1: Total number of views across languages on Wikipedia pages dedicated to 334 European cities, including 56 capitals and cities with more than 200,000 inhabitants.

across cities in Europe and the number of views in Ukrainian on Wikipedia pages dedicated to European cities. As a case study, we focus on the number of Ukrainian refugees in Poland. The data from UNHCR[5] consists of the daily number of Ukrainian refugees crossing the border from Ukraine to Poland since 24th of February 2022.

Ethical considerations

We only collect publicly available data through the UNHRC website, Wikimedia pageviews API, and MediaWiki API, following ethical guidelines (Rivers and Lewis 2014). Our study uses only aggregated data regarding the number of refugees in Poland, and views on Wikipedia pages. Regarding Wikipedia data, for privacy reasons, the geographic location of readers on a per-page basis is not available, and we do not attempt to identify users nor link any personal information to any particular user.

Case Study: Ukraine

According to the United Nations (UN), more than 5.2 million refugees from Ukraine have been recorded across Europe. The top countries used to cross the borders from Ukraine are Russia (approx. 1,412,425 Ukrainian refugees), Poland (1,194,642), Moldova (82,700), Romania (83,321), Slovakia (79,770), Hungary (25,800), and Belarus (9,820).

Since Poland is one of the top countries receiving Ukrainian refugees, our analysis focuses on the number of views on some Ukrainian Wikipedia pages dedicated to Polish cities. We compare the daily number of Ukrainian refugees crossing the border to Poland with the number of views on Wikipedia pages dedicated to Polish cities.

[5]https://data.unhcr.org/es/situations/ukraine/location/10781

Figure 2 shows the proportion of views per day on Wikipedia pages dedicated to some cities in Poland with more than 300,000 inhabitants in four different languages: English, Polish, Russian, and Ukrainian. Across all of them, we observe that the proportion of views tends to be higher in the Polish language at the beginning of the times series and stable all over the years. The proportion of views on English Wikipedia dedicated to Polish cities seems to be stable, with some non-regular picks over the years. Finally, the Ukrainian Wikipedia dedicated to those cities is the most interesting one. The proportion of views on Wikipedia pages dedicated to Polish cities written in the Russian or Ukrainian language increased dramatically in 2022, the most intense year in the war between Ukraine and Russia so far.

As a validation step, we correlate the number of views on Ukrainian Wikipedia dedicated to Polish cities with official statistics provided by the UNHCR regarding the number of refugees crossing the border from Ukraine to Poland since February 2022. Figure 3 shows the correlation between the number of Ukrainian refugees crossing the border from Ukraine to Poland (since 24 February 2022) and the number of views Wikipedia pages dedicated to some cities in Poland across different languages: English, Polish, Russian and Ukrainian. In most of the Wikipedia pages dedicated to Polish cities, the strongest correlation occurs between the number of views in the Ukrainian language and the number of refugees in Poland.

Discussion and Conclusion

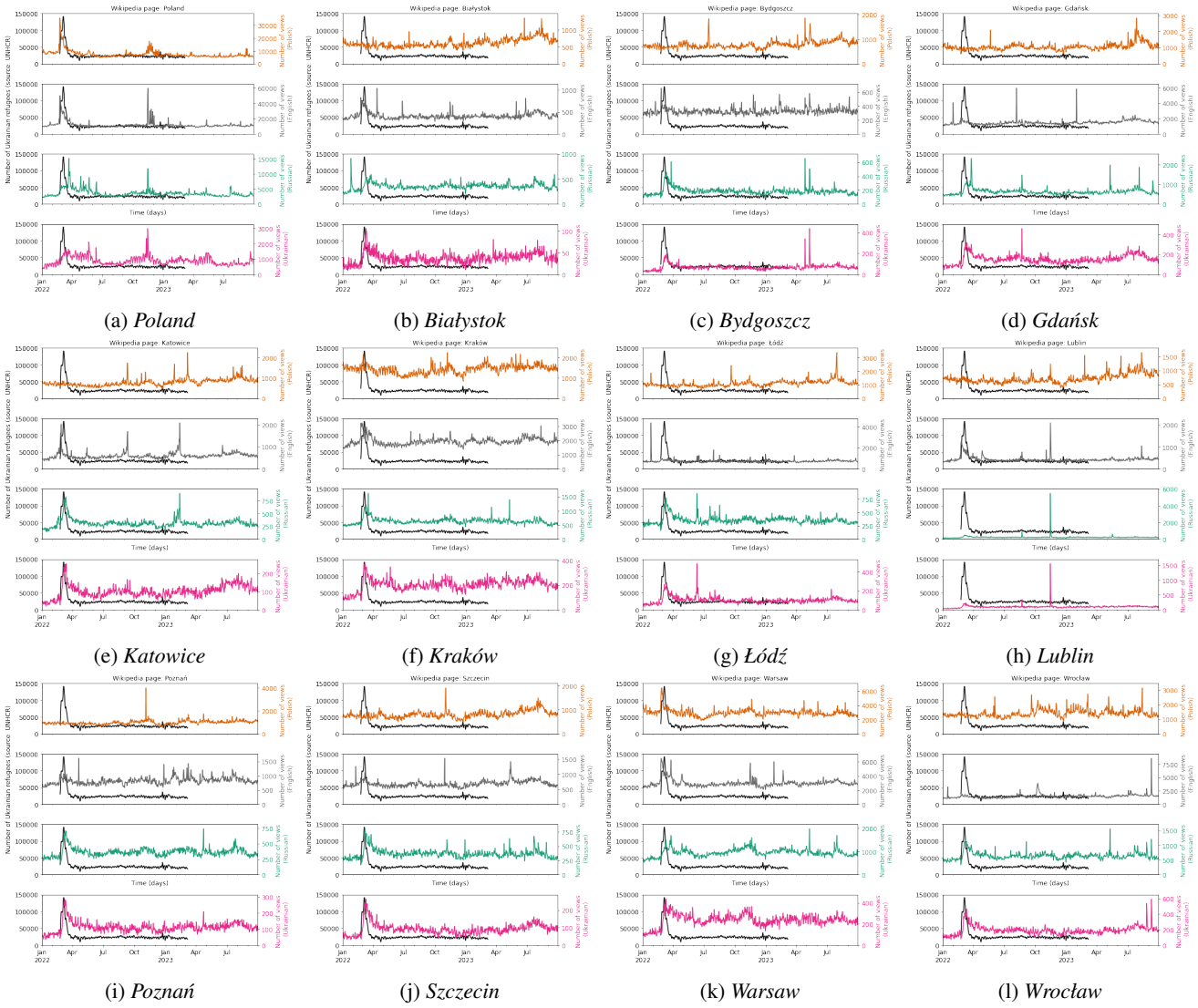
There are many challenges involved in tracking and predicting migration flows, especially when these migration events are related to unexpected reasons, such as conflicts, wars, and disasters. For instance, traditional data sources, such as surveys, require a lot of time, effort, and high costs. An alternative to tracking, predicting, and assessing real-world changes faster than traditional data sources is via digital trace data.

In this study, we propose the use of Wikipedia as a new data source to study the online response to events in the real world, especially related to migration. In our case study, we focused on the high number of Ukrainian refugees in Europe due to the Russian invasion of Ukraine in 2022. We observed that the number of views on Ukrainian Wikipedia pages dedicated to cities in Poland is in alignment with the number of Ukrainian refugees in Poland.

These results reveal opportunities in the use of Wikipedia as a proxy to study and predict mass migration flows and contribute to the literature on the relationship between information networks and migration networks.

Implications

To the best of our knowledge, our work is a first attempt to use Wikipedia views data to monitor migration events. We propose a methodology timely, cost-effective, reproducible, and scalable using Wikipedia as a new data source to monitor migration events in real-time. There are a couple of implications in this work.

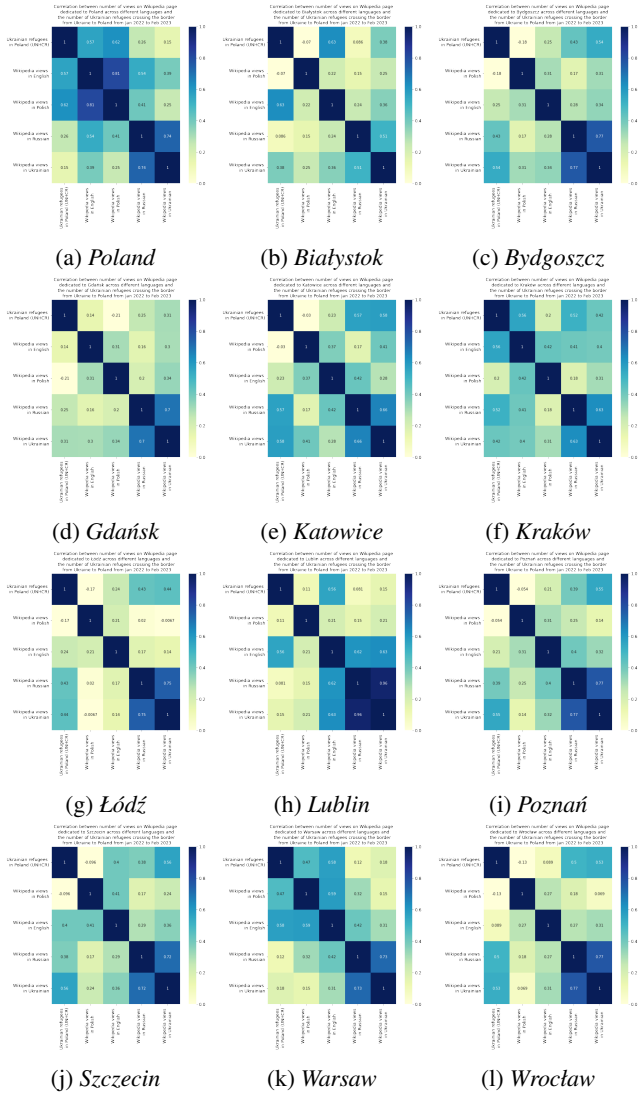


(a) *Poland* (b) *Białystok* (c) *Bydgoszcz* (d) *Gdańsk*

(e) *Katowice* (f) *Kraków* (g) *Łódź* (h) *Lublin*

(i) *Poznań* (j) *Szczecin* (k) *Warsaw* (l) *Wrocław*

Figure 2: Proportion views per day on Wikipedia pages dedicated to some cities in Poland across different languages (i.e., English as a baseline, Polish, and Ukrainian).



(a) *Poland* (b) *Białystok* (c) *Bydgoszcz*

(d) *Gdańsk* (e) *Katowice* (f) *Kraków*

(g) *Łódź* (h) *Lublin* (i) *Poznań*

(j) *Szczecin* (k) *Warsaw* (l) *Wrocław*

Figure 3: Correlation between the official number of Ukrainian refugees crossing the border from Ukraine to Poland (since 24 February 2022) and the number of views Wikipedia pages dedicated to some cities in Poland across different languages (i.e., English as a baseline, Polish, and Ukrainian).

First, from a political point of view, decision-makers would benefit from real-time estimates regarding big migration events. This implication itself is extremely important and could benefit especially countries where the political situation is unstable.

Second, a big question in demography or social sciences, in general, relates to predicting *When/Where is going to happen/be the next crisis or the next big wave of migrants?*. Similarly to the first point, our methodology could help to address this question rapidly. Moreover, the methodology could be easily adapted to other contexts and the data from Wikimedia itself could be incorporated as a new data source for social science projects in general.

Finally, our methodology repurposed the use of Wikipedia data to study an important real-world phenomenon. Given the real-world impacts of this project, Wikipedia also benefits from it.

Limitations and Future Work

Besides all the positive implications, our work has also a couple of limitations. We use language as a proxy for the country of origin. However, it is important to notice that some languages are broadly spoken in more than one country and could potentially bias the results. For instance, Russian is spoken in, at least, four countries as an official language. This limitation is not a big concern in the case of the Ukrainian language, since Ukrainian is concentrated around Ukraine. In addition to that, it is natural to think about an increase in the number of views on Wikipedia pages dedicated to countries or cities affected by real-world events in response to the event itself. For instance, we showed a strong correlation between the number of Ukrainian refugees in Poland and the number of views on Ukrainian Wikipedia dedicated to Polish cities. However, we do not establish any causal inference in this paper.

In future work, the causal inference could be better investigated, and the data collection expanded to cover more cities affected differently by the Ukrainian refugee flows. Additionally, we would like to assess the impact of edits on Wikipedia pages on refugees seeking information about the place they are planning to move to. Connectivity is also a big factor in migration and the dedication, size of the editors' community, and quality of the information provided by Wikipedia editors can work as a push factor for the migration of refugees. Finally, Wikipedia data can be used to access different stages of migration: (i) views to certain pages can indicate the number of users who are likely to move to a city or country, (ii) edits reveal the size of the community in the city or country, and finally, (iii) edits in a specific language followed by edits in other languages for that same page reveals the community connection between those languages.

References

Agarwal, P.; Redi, M.; Sastry, N.; Wood, E.; and Blick, A. 2020. Wikipedia and Westminster: Quality and Dynamics of Wikipedia Pages about UK Politicians. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, HT '20, 161–166. New York, NY, USA: Association

for Computing Machinery. ISBN 978-1-4503-7098-1. doi: 10.1145/3372923.3404817. URL https://doi.org/10.1145/3372923.3404817.

Alexander, M.; Polimis, K.; and Zagheni, E. 2019. The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data. *Population and Development Review* 617–630.

Avramescu, A.; and Wiśniowski, A. 2021. Now-casting Romanian migration into the United Kingdom by using Google Search engine data. *Demographic Research* 45: 1219–1254.

Bipat, T.; McDonald, D. W.; and Zachry, M. 2018. Do We All Talk Before We Type? Understanding Collaboration in Wikipedia Language Editions. In *Proceedings of the 14th International Symposium on Open Collaboration*, OpenSym '18, 1–11. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5936-8. doi: 10.1145/3233391.3233542. URL https://doi.org/10.1145/3233391.3233542.

Böhme, M. H.; Gröger, A.; and Stöhr, T. 2020. Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics* 142: 102347.

Cesare, N.; Lee, H.; McCormick, T.; Spiro, E.; and Zagheni, E. 2018. Promises and pitfalls of using digital traces for demographic research. *Demography* 55(5): 1979–1999.

Eom, Y.-H.; Aragón, P.; Laniado, D.; Kaltenbrunner, A.; Vigna, S.; and Shepelyansky, D. L. 2015. Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions. *PLOS ONE* 10(3): e0114825. ISSN 1932-6203. doi:10.1371/journal.pone.0114825. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114825. Publisher: Public Library of Science.

Golenvaux, N.; Alvarez, P. G.; Kiossou, H. S.; and Schaus, P. 2020. An LSTM approach to Forecast Migration using Google Trends. *arXiv preprint arXiv:2005.09902* .

Graham, M.; and Dittus, M. 2022. *Geographies of Digital Exclusion: Data and Inequality*. JSTOR.

Kim, S.; Park, S.; Hale, S. A.; Kim, S.; Byun, J.; and Oh, A. H. 2016. Understanding Editing Behaviors in Multilingual Wikipedia. *PLOS ONE* 11(5): e0155305. ISSN 1932-6203. doi:10.1371/journal.pone.0155305. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155305. Publisher: Public Library of Science.

Lehmann, J.; Müller-Birn, C.; Laniado, D.; Lalmas, M.; and Kaltenbrunner, A. 2014. Reader preferences and behavior on Wikipedia. In *Proceedings of the 25th ACM conference on Hypertext and social media*, 88–97. Santiago Chile: ACM. ISBN 978-1-4503-2954-5. doi:10.1145/2631775.2631805. URL https://dl.acm.org/doi/10.1145/2631775.2631805.

Lerner, J.; and Lomi, A. 2019. Let's Talk About Refugees: Network Effects Drive Contributor Attention to Wikipedia Articles About Migration-Related Topics. In Aiello, L. M.; Cherifi, C.; Cherifi, H.; Lambiotte, R.; Lió, P.; and Rocha, L. M., eds., *Complex Networks and Their Applications VII*, Studies in Computational Intelligence, 211–222. Cham: Springer International Publishing. ISBN 978-3-030-05414-4. doi:10.1007/978-3-030-05414-4_17.

Leurs, K.; and Smets, K. 2018. Five questions for digital migration studies: Learning from digital connectivity and forced migration in (to) Europe. *Social Media+ Society* 4(1): 2056305118764425.

Lin, A. Y.; Cranshaw, J.; and Counts, S. 2019. Forecasting us domestic migration using internet search queries. In *The world wide web conference*, 1061–1072.

Lucchini, L.; Tonelli, S.; and Lepri, B. 2019. Following the footsteps of giants: modeling the mobility of historically notable individuals using Wikipedia. *EPJ Data Science* 8(1): 36. ISSN 2193-1127. doi:10.1140/epjds/s13688-019-0215-7. URL https://epjds.epj.org/articles/epjdata/abs/2019/01/13688_2019_Article_215/13688_2019_Article_215.html. Number: 1 Publisher: Springer Berlin Heidelberg.

Miquel-Ribé, M.; and Laniado, D. 2019. Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions. *Proceedings of the International AAAI Conference on Web and Social Media* 13: 620–629. ISSN 2334-0770. doi:10.1609/icwsm.v13i01.3260. URL https://ojs.aaai.org/index.php/ICWSM/article/view/3260.

Miquel-Ribé, M.; and Laniado, D. 2020. The Wikipedia Diversity Observatory: A Project to Identify and Bridge Content Gaps in Wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration*, OpenSym '20, 1–4. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8779-8. doi:10.1145/3412569.3412866. URL https://doi.org/10.1145/3412569.3412866.

Moat, H. S.; Curme, C.; Avakian, A.; Kenett, D. Y.; Stanley, H. E.; and Preis, T. 2013. Quantifying Wikipedia Usage Patterns Before Stock Market Moves. *Scientific Reports* 3(1): 1801. ISSN 2045-2322. doi:10.1038/srep01801. URL http://www.nature.com/articles/srep01801.

Panciera, K.; Halfaker, A.; and Terveen, L. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedinfs of the ACM 2009 international conference on Supporting group work - GROUP '09*, 51. Sanibel Island, Florida, USA: ACM Press. ISBN 978-1-60558-500-0. doi:10.1145/1531674.1531682. URL http://portal.acm.org/citation.cfm?doid=1531674.1531682.

Perrotta, D.; Johnson, S. C.; Theile, T.; Grow, A.; de Valk, H.; and Zagheni, E. 2022. Openness to migrate internationally for a job: evidence from LinkedIn data in Europe. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, 759–769.

Ribeiro, M. H.; Gligorić, K.; Peyrard, M.; Lemmerich, F.; Strohmaier, M.; and West, R. ???? Sudden Attention Shifts on Wikipedia During the COVID-19 Crisis 12.

Rivers, C. M.; and Lewis, B. L. 2014. Ethical research standards in a world of big data. *F1000Research* 3(38): 38.

Ruprechter, T.; Horta Ribeiro, M.; Santos, T.; Lemmerich, F.; Strohmaier, M.; West, R.; and Helic, D. 2021. Volunteer contributions to Wikipedia increased during COVID-19 mobility restrictions. *Scientific Reports* 11(1): 21505. ISSN 2045-2322. doi:10.1038/s41598-021-00789-3. URL https://www.nature.com/articles/s41598-021-00789-3. Number: 1 Publisher: Nature Publishing Group.

Singer, P.; Lemmerich, F.; West, R.; Zia, L.; Wulczyn, E.; Strohmaier, M.; and Leskovec, J. 2017. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, 1591–1600. Perth Australia: International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4913-0. doi:10.1145/3038912.3052716. URL https://dl.acm.org/doi/10.1145/3038912.3052716.

Spyratos, S.; Vespe, M.; Natale, F.; Weber, I.; Zagheni, E.; and Rango, M. 2018. Migration data using social media: a European perspective .

The GSM Association. 2017. The Importance of Mobile for Refugees: A Landscape of New Services and Approaches. URL: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/02/The-Importance-of-mobile-for-refugees_a-landscape-of-new-services-and-approaches.pdf. Online; accessed Apr. 2022.

Tjaden, J. 2021. Measuring migration 2.0: a review of digital data sources. *Comparative Migration Studies* 9(1): 1–20.

Twyman, M.; Keegan, B. C.; and Shaw, A. 2017. Black Lives Matter in Wikipedia: Collective Memory and Collaboration around Online Social Movements. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, 1400–1412. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-4335-0. doi:10.1145/2998181.2998232. URL https://doi.org/10.1145/2998181.2998232.

Ulutürk, I.; Uysal, I.; and Varol, O. 2019. Refugee integration in Turkey: a study of mobile phone data for D4R challenge. In *Data for refugees challenge workshop*.

Vieira, C. C.; Fatehkia, M.; Garimella, K.; Weber, I.; and Zagheni, E. 2022. Using Facebook and LinkedIn Data to Study International Mobility .

Vilain, P.; Larrieu, S.; Cossin, S.; Caserio-Schönemann, C.; and Filleul, L. 2017. Wikipedia: a tool to monitor seasonal diseases trends? *Online Journal of Public Health Informatics* 9(1): e052. ISSN 1947-2579. doi:10.5210/ojphi.v9i1.7630. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5462208/.

West, R.; Weber, I.; and Castillo, C. 2012. A data-driven sketch of Wikipedia editors. In *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion*, 631. Lyon, France: ACM Press. ISBN 978-1-4503-1230-1. doi:10.1145/2187980.2188162. URL http://dl.acm.org/citation.cfm?doid=2187980.2188162.

Yoshida, M.; Arase, Y.; Tsunoda, T.; and Yamamoto, M. 2015. Wikipedia Page View Reflects Web Search Trend. In *Proceedings of the ACM Web Science Conference*, 1–2. Oxford United Kingdom: ACM. ISBN 978-1-4503-3672-7. doi:10.1145/2786451.2786495. URL https://dl.acm.org/doi/10.1145/2786451.2786495.

Zhu, T. J.; Fritzler, A.; and Orlowski, J. A. K. 2018. World Bank Group-LinkedIn Data Insights: Jobs, Skills and Migration Trends Methodology and Validation Results .