

Using Wikipedia data to monitor unexpected migration events

Migration, Refugees, Ukraine, Syria, Wikipedia

Extended Abstract

Migration flows are challenging in terms of monitoring and prediction. The challenge increases when these migration events are unexpected, for instance, related to forced migration [3, 6]. Digital trace data, such as Google Trends [1], Facebook [5], and LinkedIn Ads [8], have been shown as good data sources to monitor and predict migration. In fact, digital trace data reflect changes in the real world faster than traditional data sources. In this work, we explore data from Wikipedia to monitor and provide an early warning for unexpected mass migration events.

Wikipedia is the largest and most popular free online encyclopedia aiming to provide access to information about current events and media coverage of a topic worldwide [4]. Wikipedia pages are created and edited by volunteers around the world [2]. 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 family¹ 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 use Wikipedia data 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 [7]. Wikipedia is a worldwide well-known source of information. By examining the association between information changes (i.e., edits) and access (i.e., views) on Wikipedia by language and recent refugee flows, we aim to determine whether Wikipedia data can be used to monitor and predict mass migration flows.

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 access to information on Wikipedia and mass migration events, thus showing how it can be used as a predictor for ongoing mass and unexpected migration flows as well as a broader proxy of early warning for policymakers. Wikipedia provides absolute numbers of views and edits, which is an advantage compared to other sources of information, such as Google Trends.

Our methodology consists of using Wikipedia data to assess how the number of views and edits on Wikipedia pages dedicated to cities around the world change across time in response to migration events. As a case study, we focus on two recent migration events caused by wars. The first one refers to refugees of the Syrian civil war in 2011 and the second one to the Ukrainian refugee crisis in 2022. For each Wikipedia page dedicated to the capitals and most populous European cities, we investigate the number of edits in Arabic (i.e., the primary language in Syria) and Ukrainian (i.e., the official language in Ukraine). We observed that the number of edits on Wikipedia pages dedicated to European countries in the Arabic language increased around 2012, corresponding to the period when Europe started receiving a lot of

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

Syrian refugees. Similarly, especially on Wikipedia pages dedicated to countries that share borders with Ukraine, the number of edits in Ukrainian increased in 2016 and 2022. In fact, 2014-2016 followed by 2022 correspond to periods of time where the number of Ukrainian refugees in Europe increased the most. This result confirms that information regarding changes in the real world is rapidly incorporated into Wikipedia pages.

Next, we analyzed the number of views on Wikipedia pages dedicated to cities as a proxy for migration intentions to those cities. We collected the number of views on Ukrainian Wikipedia pages dedicated to the most populous cities in Poland. According to the data provided by the United Nations High Commissioner for Refugees (UNHCR), Poland is the country with more Ukrainian refugees. We observed an increase in the number of views on the Ukrainian Wikipedia dedicated to Polish cities in 2022 (see Figure 1). In fact, the increasing number of views started around the beginning of February, even before the war started on February 24th, 2022. Additionally, we observed a strong association² between the number of Ukrainian refugees crossing the border from Ukraine to cities in Poland (according to the UNHCR) and the number of views on Ukrainian Wikipedia dedicated to Polish cities. These results indicate that Wikipedia views are a good predictor for migration flows.

Finally, we contribute to the literature on the use of digital trace data to study unexpected migration events in real-time. Moreover, we contribute to the literature on the relationship between information networks and migration networks by showing that edits on Wikipedia pages reflect changes in the real world (e.g., big migration events). Also, our results reveal opportunities in the use of Wikipedia as a proxy to study and predict mass migration flows.

References

- [1] M H Böhme, A Gröger, and T Stöhr. Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 2020.
- [2] M Graham and M Dittus. *Geographies of Digital Exclusion: Data and Inequality*. JSTOR, 2022.
- [3] K Leurs and K Smets. Five questions for digital migration studies: Learning from digital connectivity and forced migration in (to) europe. *Social Media+ Society*, 4(1):2056305118764425, 2018.
- [4] P Singer, F Lemmerich, R West, L Zia, E Wulczyn, M Strohmaier, and J Leskovec. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1591–1600, April 2017.
- [5] S Spyrtatos, M Vespe, F Natale, I Weber, E Zagheni, and M Rango. Migration data using social media: a european perspective. 2018.
- [6] J Tjaden. Measuring migration 2.0: a review of digital data sources. *Comparative Migration Studies*, 9(1):1–20, 2021.
- [7] I Ulutürk, I Uysal, and O Varol. Refugee integration in turkey: a study of mobile phone data for D4R challenge. In *Data for refugees challenge workshop*, 2019.
- [8] C C Vieira, M Fatehkia, K Garimella, I Weber, and E Zagheni. Using facebook and linkedin data to study international mobility. 2022.

²The correlation varies from 0.11 to 0.61.

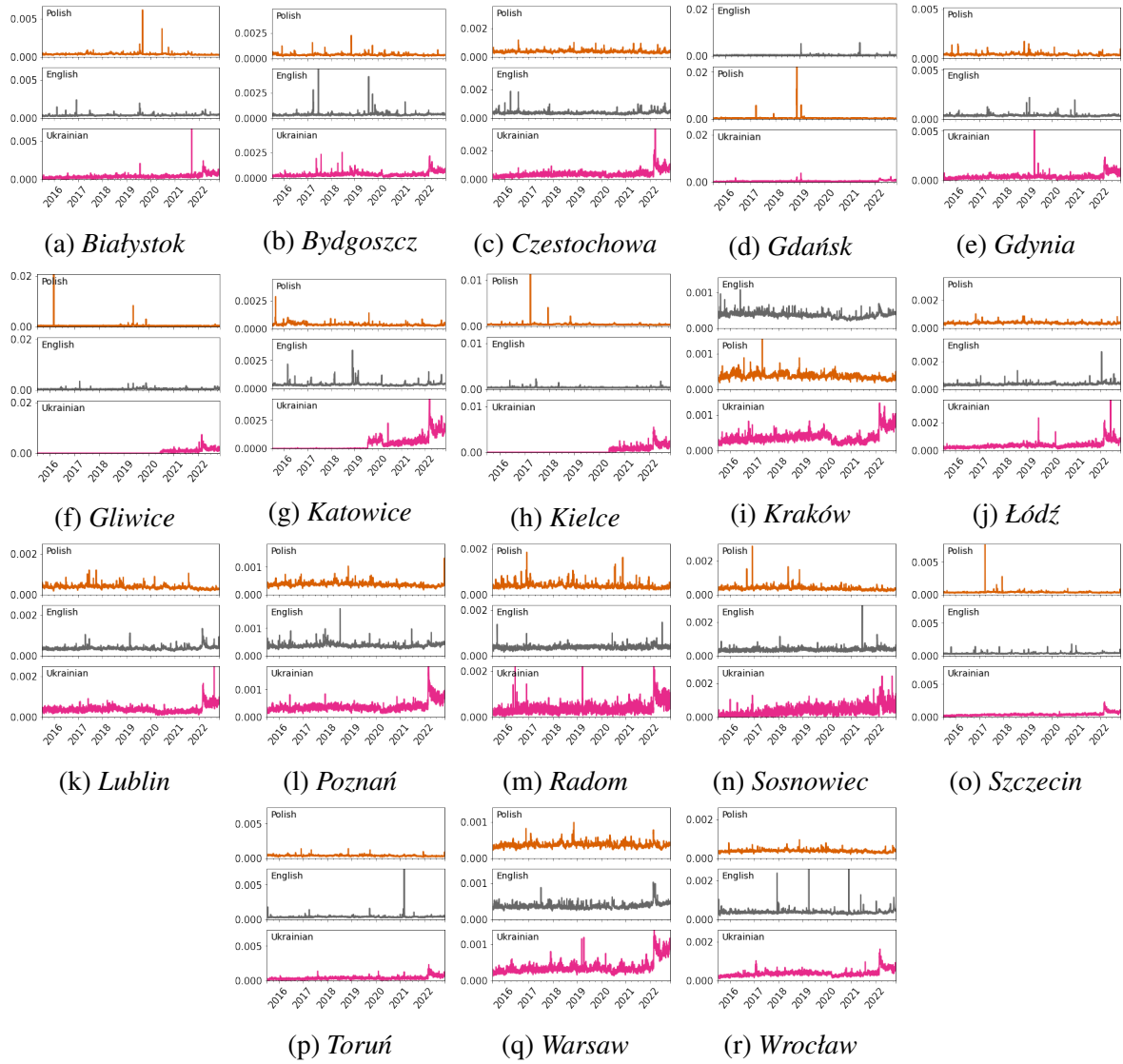


Figure 1: Proportion views per day on Wikipedia pages dedicated to some cities in Poland across different languages (i.e., English as a baseline, Polish, and Ukrainian).