HiDALGO

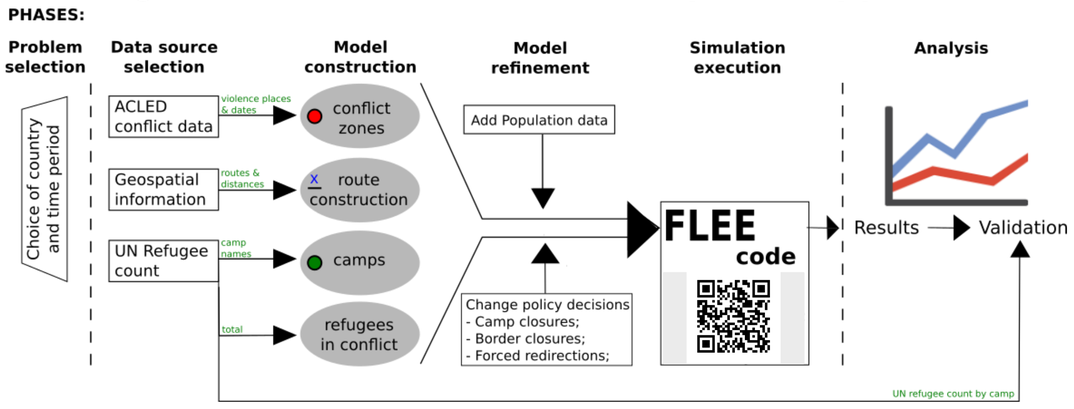
AI-Support for large-scale Refugee Movement Simulations

Martin Maritsch[1], Diana Suleimenova[2], Bernhard C. Geiger[1], Derek Groen[2]
[1]Know-Center GmbH, Graz, Austria; [2]Department of Computer Science, Brunel University, London, UK

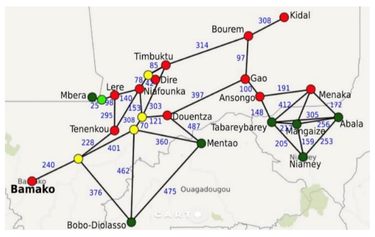
Agent-Based Simulation

Agent-based simulations (ABS) are based on interacting autonomous agents. We use ABS to perform **refugee movement simulations**, and have validated our approach (approx. **75% accuracy** in destinations) against **three African conflicts**[1,2]. To **improve accuracy** and use our approach for a wider range of research questions, we need to clarify the values of several **key parameters** through **big data analytics**.

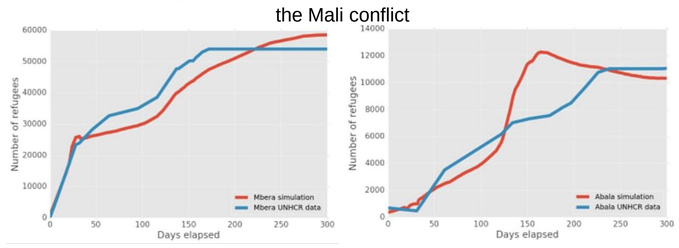
A generalised simulation development approach



Overview of geographic network model for Mali



Number of refugees as predicted by our simulation and obtained from the UNHCR data for the Mali conflict



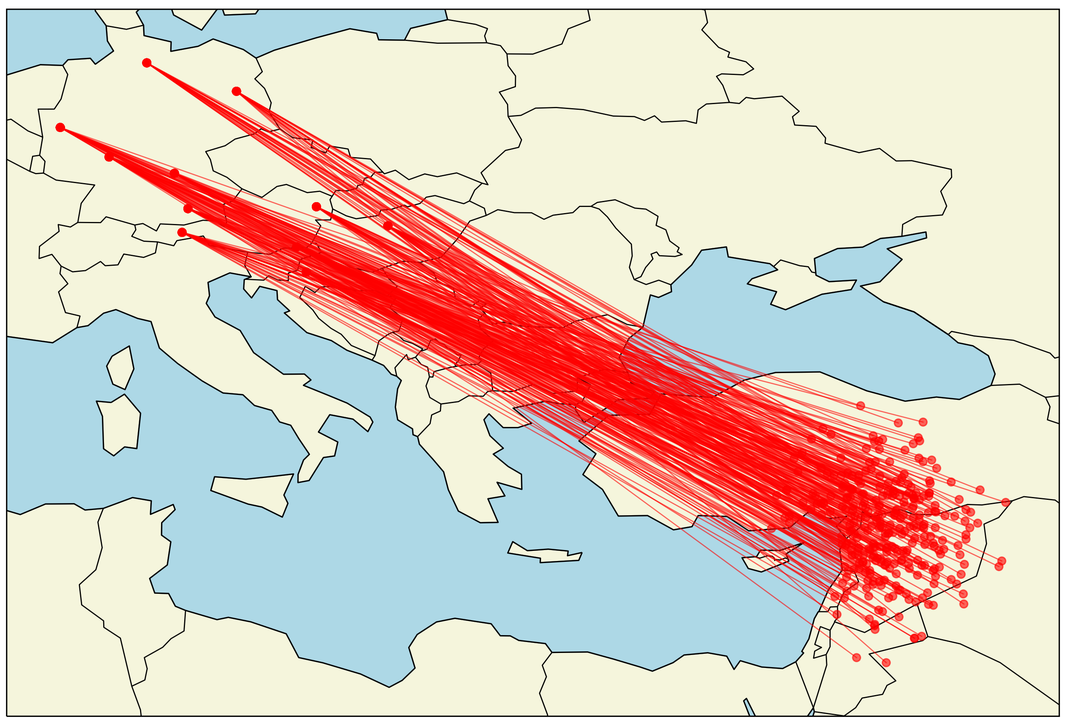
[1] Suleimenova D, Bell, D. & Groen D. (2017) “A generalized simulation development approach for predicting refugee destinations”, S*cientific Reports*, 7:13377.
[2] Suleimenova D., Bell D. & Groen D. (2017) “Towards an automated framework for agent-based simulation of refugee movements”, in W. K. V. Chan, A. DAmbrogio, G. Zacharewicz, N. Mustafee, G. Wainer and E. Page (eds.), *Proceedings of the 2017 Winter Simulation Conference*, (pp. 1240–1251). Las Vegas, Nevada: IEEE.

Data & Features

Call Detail Records (CDRs)

CDRs may be collected by different mobile network carriers and roaming providers and hence may differ in their attributes and accuracy. Commonly they at least contain **caller/callee information** (phone number and location) and a **timestamp**. For our use case, we will primarily leverage international CDRs from call routers spanning over multiple countries since we are interested in refugees placing calls to their destination and source country while on the move.

Simulated phone calls from the Syria crisis region to destinations in central Europe



Feature Engineering

One of the key tasks is to **classify** which records in the available CDRs correspond to **refugees and which do not**. We can construct and utilize the following **features** from the dataset:

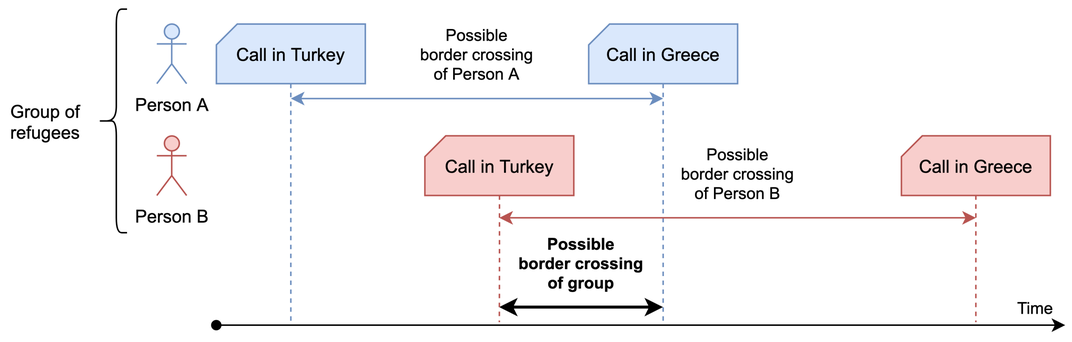
- Multiple border crossings
- One-way journey
- Duration of stay in each country
- Calls made to relevant countries
- Mode of transportation (determined by travel speed)
- Number of countries visited

By utilizing **additional data sources**, such as **official refugee counts** from UNHCR or **conflict data** from ACLED we can **validate** and **extend** our evolved models.

AI for Understanding Movement Patterns

Once we have **pre-processed the dataset** and **extracted relevant features**, we will leverage **AI methods** to extract knowledge from the data. Using the features mentioned before, we can establish models of **how and in what patterns** groups of refugees move. Further we expect to find out, how the **group size influences** the **behaviour of people** (e.g., they travel faster or slower or make fewer or more stops when in bigger groups).

The knowledge about refugee groups can be intelligently used to improve the understanding of refugee movements



When? Where To? How Many?

Leveraging real-time CDRs we anticipate to be able to **predict refugee movement outbreaks** based on patterns in historical data. We assume that people's plans to move from a country reflect in their calling behaviour, specifically we expect that they might place more calls to possible destination countries. In this setting, the training target is set to be the number of caller identifiers leaving the country to a given destination in the following week or month. Furthermore, we will bring CDRs into context with conflict data from ACLED. Recognizing patterns in historical CDRs in events such as battles, riots or protests can help to timely identify the occurrence of such events based on analysis of real-time CDRs.

Destination countries of refugee movements can be estimated by using a probabilistic model based on the observed CDRs. Trivially, we can assign countries probabilities of being a potential destination for refugees (e.g., the probability for Germany being a destination would be rather high as opposed to Syria). Important factors are the per-country distribution of relevant CDRs.

In terms of estimating the **number of people** expected to leave, we hypothesize that this number can be predicted given the amount of calls combined with a constant factor representing the penetration of mobile phones for each country. Additionally, linking CDRs with official refugee counts from UNHCR allows us to correlate call behaviour (e.g., quantity of calls made to a destination country) with actual refugee counts. These observations are then useful for future predictions.

AI-supported Refugee Movement Simulations

From the extracted knowledge, we will provide fundamental **simulation parameters** and agent rule-sets to be used in the migration **ABS** investigated in *HiDALGO*, e.g.:

- Average **movement speed**
- **Movement probability** for location types (camps, cities, conflict areas)
- People behaviour in **group formation**

Furthermore, using real-time CDRs we input parameters to ongoing simulations:

- Estimated **refugee counts** based on calling behaviour
- **Starting location** and **size** of **refugee movements**
- Potential **destination countries**

Architectural Implementation

The mining of CDRs can be modelled as a classical batch job in an ETL-pipeline since the nature of the data allows us to **highly parallelize** processing. Within *HiDALGO*, we will use **Apache Spark** on high-performance infrastructures. There, we will make use of MPI extensions (e.g. **Alchemist**).

Machine Learning tasks will be implemented with Spark’s **MLlib** and extensions for Alchemist. For the real-time analysis **Spark Streaming** will be utilized.

Acknowledgments

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

Brunel University: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement nos. 800925 and 671564.

https://hidalgo-project.eu

contact@hidalgo-project.eu

Coordinator:

ATOS SPAIN
Calle de Albarracin 25
Madrid 28037
SPAIN

Project #824115 funded by the Horizon 2020 Framework Programme of the European Union.