



Shared Prosperity Dignified Life



DATA-POP
ALLIANCE

Changing the world with data

Assessing the living conditions of refugee communities in Turkey and Lebanon through the lens of cell phone metadata: promise, challenges, and future directions

Beirut, Lebanon

December 23, 2020

Outline

- Context
- Case Study: Turkey
 - Data for refugees challenge
 - Key Findings
- Challenges
- Limitations
- Ethical and Privacy Concerns and Solutions



Context



Context

- After the start of the Syrian Civil War in 2011–12, increasing numbers of civilians sought refuge in neighboring countries.
- By May 2017, **Turkey had received over 3 million refugees** (the largest refugee population in the world).
- Problems: integration, income, welfare, employment, health, education, language, social tension, and discrimination.
- A good understanding of refugee dynamics is necessary.



Case Study: Turkey



Turkey Example

D4R Challenge

- Initiated to improve the conditions of Syrian refugees in Turkey
- Anonymized CDRs of calls and SMS from Türk Telekom
- Research topics included health, education, unemployment, safety, and social integration concerns



CDR Data

- 992,457 customers of Türk Telekom, of which 184,949 are tagged as “refugees”, and 807,508 as Turkish citizens
- 75% of refugee-tagged customers are recorded as “male” and 25% as “female”

Data Structure

- site-to-site hourly traffic, aggregated to site level
- Anonymized user-level 2-week activity with tower-cell location
- Anonymized user-level activity with district location



Key Findings from Turkey

Labor - Refugee Mobility

- One in five Syrian refugees moved to at least one other province within a single year
- Access to rich provinces is limited for refugees
- Determinants of mobility: income of the origin/source province, distance between provinces
- Refugees move more frequently and in smaller areas



Labor - Seasonal Work

- Increased CDR activity in certain regions and certain months of the year
- **E.g.** growth rate of average weekly number of refugee-tagged calls in Ordu compared to periods before, during and after harvest season
- Similar results for touristic regions and seasons (e.g. Antalya)



Labor - Detecting Employment

- Regularity of individual commuting patterns
 - Work-Time and Home-Time Locations
- Mobile calls made from specific tower repeatedly during the home time (e.g. weekday evenings, weekends)
- Mobile calls made from different tower repeatedly during the work time (e.g. weekdays 9-5)
- Clustering algorithms and similarity matrices exist to do this
- Based on geographic zone of tower (e.g. touristic and leisure places)



Safety and Security

- Triangulation of CDRs, social media sentiment, and refugee camp locations in reference to qualitatively relevant violent events
 - violent events tend to occur more in places with negative sentiments and high refugee concentrations
 - Real-time monitoring can be used for timely interventions to diffuse tensions in the local population
- Planning for Disaster Scenarios
 - CDR based concentration to optimize emergency response in natural disaster scenarios



Social Integration

- Integration typically measured with qualitative approaches
- CDR enabled assessment of social, spatial, and economic integration
 - **E.g.** calls from refugees to natives and vice versa
- Assessing integration in an area needs to be complemented with local indicators
 - labor market, the real estate market, and social tension indicators



Other Applications

Education

- CDR reflect the distribution of refugees and natives
- Improve refugee access to schools
 - In some districts, average distance to the closest school is large (>5Km)
 - affordable and safe transportation for refugees needed

Health

- Estimating refugees' home addresses using CDR
- Calculation of distance from refugee homes to existing migrant health centers
- Calculating optimal locations for such centers and comparing them with existing centers locations



Challenges, Limitations and Ethical Concerns



Challenges

- Political will and strong partnerships with stakeholders
- Capacities for systematic and safe access and analysis of big data sources
- Appropriateness and soundness of the research and assumptions
- Awareness about the value added of this data
- Cooperation with telecom companies and introducing them to their benefit in social good



Limitations

- Big data analysis risks leaving out some populations
- Data-invisible individuals may be more vulnerable to exploitation
- Helpful to complement regular data collection methods
- Targeted provision for government service providers and NGOs
- Each data source comes with a different type of bias
 - need to be taken into account for correct interpretation
- Children absent from such data



Ethical and Privacy Issues

- Respect applicable laws for data processing and privacy protection
- Existing rules appear ill-suited for technological advances in this context
- Need to ensure that the databases cannot be used for surveillance or tracking of individuals
- Should not be able to track down individuals (individuals should remain anonymous)



Ethical and Privacy Solutions

- Names, real phone numbers, ID numbers, or other identifying information should be excluded
- Anonymization of the identifiers for each user
- Pseudo-random numbers representing customers shall not stored anywhere along with actual phone numbers
- Geographic and temporal data aggregation
- On-site data processing





DATA-POP
ALLIANCE

Thank you!



[@datapopalliance](https://twitter.com/datapopalliance)

contact@datapopalliance.org

www.datapopalliance.org