Big Data and Labour Migration

Tuba Bircan

Interface Demography – Vrije Universiteit Brussel

27/10/2020

OWiM - Obserwatorium Wielokulturowości i Migracji

Migration Data Sources

- Population censuses
  - Migration flows
  - Migrant stocks
- Administrative sources
  - Population registers
  - Social security databases
- Large-scale surveys
  - EU-SILC, LFS, etc.
  - Gallup World Poll

Migration Measures

- Immigration-emigration
  - Inflows –Outflows
  - Granted visas
    - Residence permits
    - Work permits
    - Study visa
    - Family reunification
  - Asylum applications
- Migrant stocks
  - Country of birth
  - Country of nationality

Gaps in Migration Statistics

- Gaps in migration stastistics (Bircan et al., 2020)
  - Definitions and typologies
  - Drivers
  - Demography/gender/hidden populations
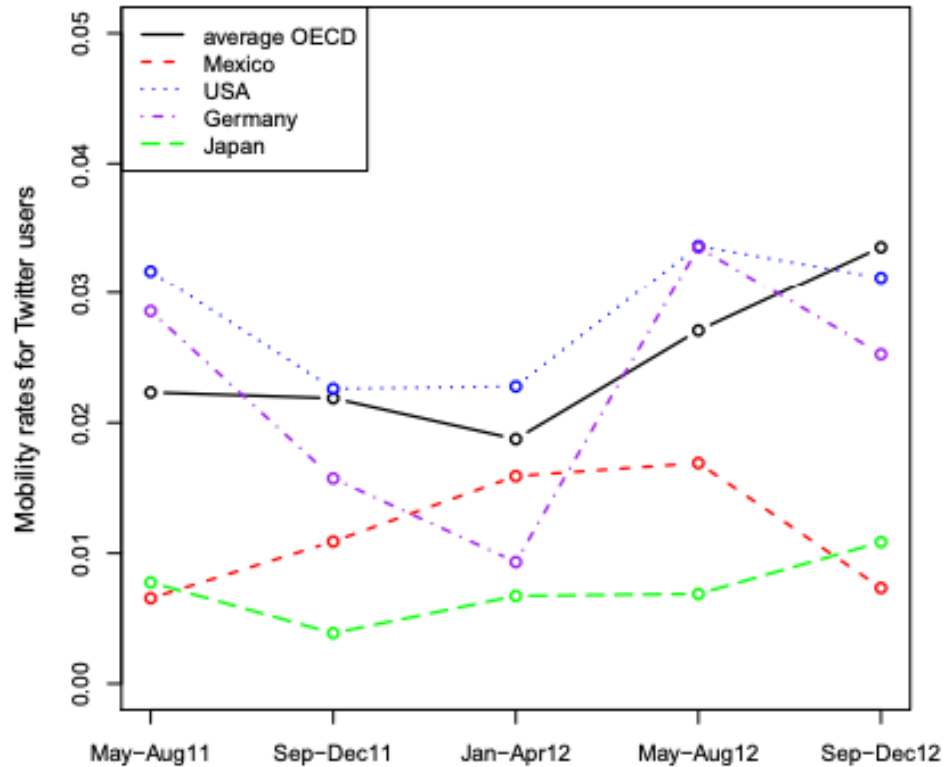 - Geography/space
  - Temporality/timeliness

Can **Big Data** Improve Migration Statistics?

- Big Data
  - Mobile-phone-based - e.g. call records or mobile money transfers.
    - CDR, XDR, etc.
  - Internet-based - e.g. social media or use of search engines.
    - Twitter (geo-coded), Facebook (API), Instagram, Google, etc.
  - Sensor-based - e.g. Earth Observation Data (satellite imagery).
    - Sentinel 1-2, etc.
  - Financial big data – e.g. banking data, remittances, etc.
  - Qualitative big data – e.g. text archives, etc.

Big Data and Migration Statistics – Social Media Data

▪ Inferring International and Internal Migration Patterns from Twitter Data (Zagheni et al., 2014)

  ▪ geolocated data for about 500,000 users of "Twitter".

    ▪ users in OECD countries during the period May 2011- April 2013.

    ▪ the geographic movements within and between countries for independent periods of four months, for the subsample of users who have posted geolocated tweets regularly

  ▪ a difference-indifferences approach to reduce selection bias when we infer trends in out-migration rates for single countries
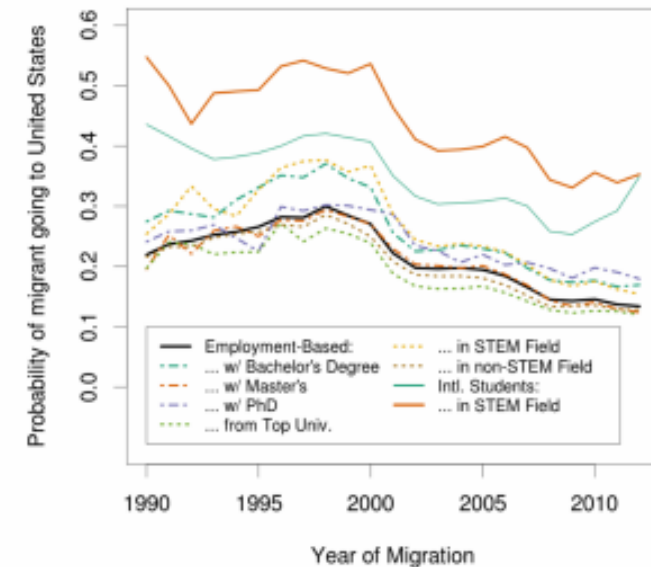
Big Data and Migration Statistics - Social Media Data

Inferring International and Internal Migration Patterns from Twitter Data (Zagheni et al., 2014)

- geolocated data for about 500,000 users of "Twitter".
  - users in OECD countries during the period May 2011-April 2013.
  - the geographic movements within and between countries for independent periods of four months, for the subsample of users who have posted geolocated tweets regularly
- a difference-indifferences approach to reduce selection bias when we infer trends in out-migration rates for single countries

Big Data and Migration Statistics - Social Media Data

- Migration of Professionals to the U.S. Evidence from LinkedIn data (State et al., 2014)

  - Geolocated career histories provided by LinkedIn for estimating the international migration of professional workers


Fig. 1. Conditional Probability of Migration to United States by Year, 1990-2012.

Big Data and Migration Statistics - Social Media Data

■ Leveraging Facebook's advertising platform to monitor stocks of migrants (Zagheni et al., 2017)

FIGURE 1   Relationship between Facebook estimates of the fraction of expats in US states (2016), by country of origin, and the respective estimates from the 2014 American Community Survey
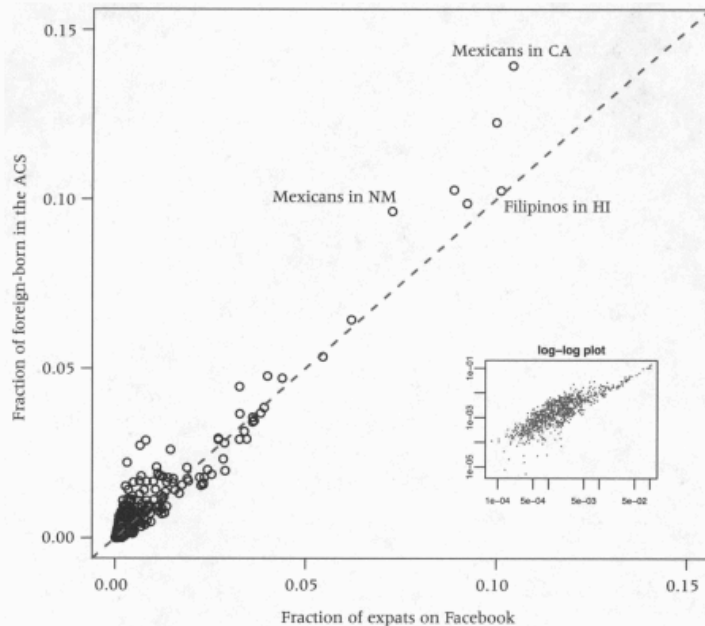
NOTES: The plot includes state/country pairs where the number of Facebook expats exceeds 1,000. The inset shows the same observations on a log-log scale. The dashed line is a 45-degree line.

■ Facebook API
  ■ socio-demographic characteristics of Facebook users, aggregated at var ious levels of geographic granularity
■ Stocks of "expats" in the USA were estimated using Facebook advertising data, and compared with the American Community Survey (ACS) to account for biases in the Facebook data

Big Data and Migration Statistics - Social Media Data

- Measuring Labour Mobility and Migration Using Big Data (Gendronneau et al, 2019)
- The aim of this study was to investigate the potential of geo-referenced social-media data to facilitate "nowcasting" stocks of EU movers and mobility flows, providing more recent estimates than official statistics to serve as early warning signs for the European Commission.

Measuring Labour Mobility and Migration Using Big Data

Exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers

Cloé Gendronneau, Arkadiusz Wiśniowski, Dilek Yildiz, Emilio Zagheni, Lee Fiorio, Yuan Hsiao, Martin Stepanek, Ingmar Weber, Guy Abel, Stijn Hoorens

Big Data and Migration Statistics - Social Media Data

- Measuring Labour Mobility and Migration Using Big Data (Gendronneau et al, 2019)

  - Collection of "big" social-media data, in particular from Facebook and Twitter, that can be used for measuring EU mobility;

  - Specification and development of a method of measurement, combining these data with traditional data sources;

  - Production of estimates for stocks of EU movers and EU mobility by EU Member State; and

  - Comparison of estimates obtained from this research with official statistics and assessment of the reliability of the method.

Big Data and Migration Statistics – Internet-based Data

▪ Migration Intentions (Böhme, Gröger and Stöhr, 2020)
- Google Trends Index (GTI) for migration-related search terms can be exploited to measure migration intentions from a certain country and predict subsequent emigration
  - The GTI data consists of high-frequency time series capturing the relative search intensities for any keyword performed through the Google search engine across the globe.
  - The GTI is by far the most representative data source for online searches worldwide with Google having a market share of more than 80% on desktop devices. This figure increases to 97% once considering mobile and tablet devices.
- Online search words
  - a set of close to fifty words specific to the migrant population, such as passport, asylum, embassy, social assistance,
  - GTI indicators for each individual keyword in the official language of the respective country of origin

Big Data and Migration Statistics - Internet-based Data

- Migration Intentions (Böhme, Gröger and Stöhr, 2020).
  - Measurement of migration intentions with consistent and representative indicators that are freely available at close to universal geographic coverage
  - Short term predictions of current migration flows ahead of official data release lags, which amount up to several years
  - Improve the performance of conventional models of the determinants of migration flows

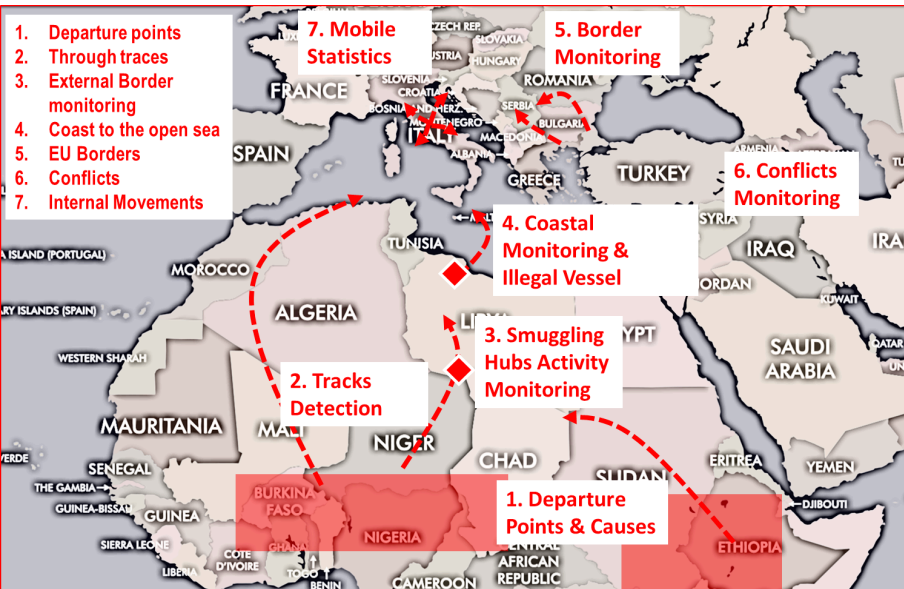# Big Data and Migration Statistics - Internet-based Data
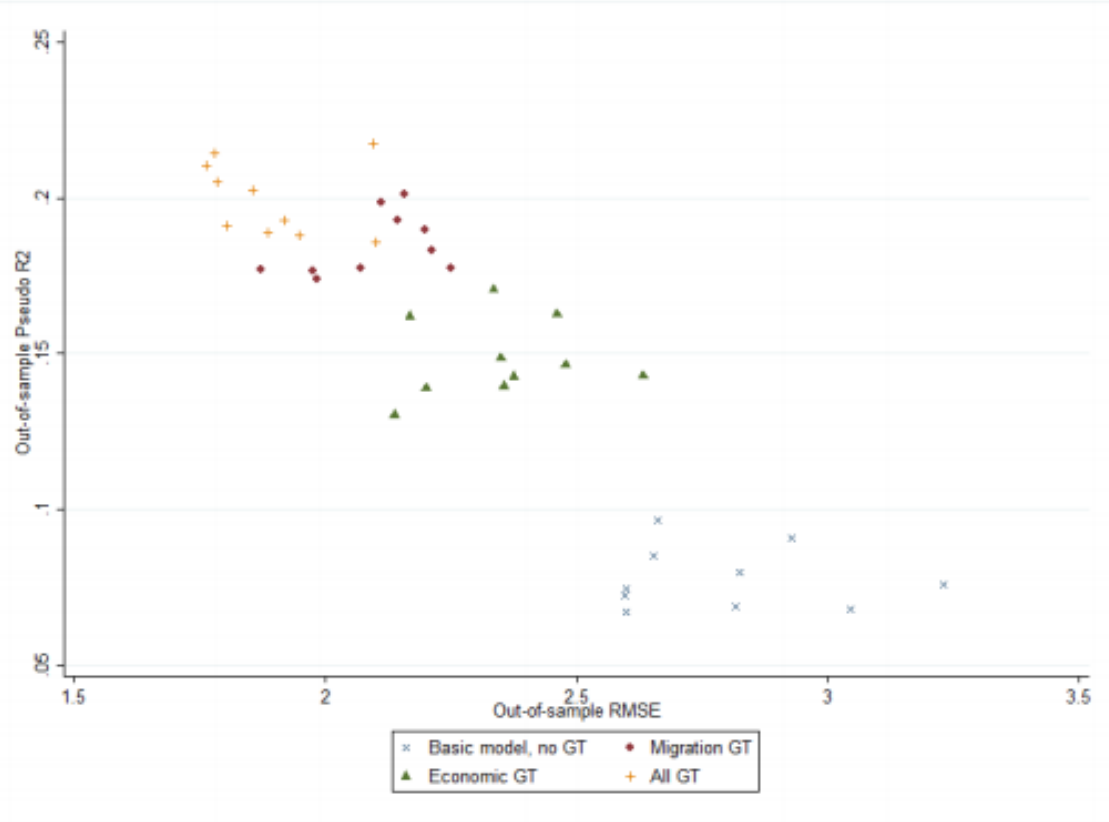
Notes: The figure reports out-of-sample estimates from 10-fold cross-validation. There are thus 10 estimates of OOS-R^2 and OOS-RMSE for each model. The basic model contains controls for log GDP and log population and origin as well as year dummies. In addition, migration keywords, economic keywords or both are added in the respective models.

- Migration Intentions (Böhme, Gröger and Stöhr, 2020).
  - improvements in the goodness of fit our model achieves in the within dimension are not due to in-sample overfit, but also holds out-of-sample.
  - The models with GTI included perform better, on average, than the basic one, explaining more of the variation in the migration outcome measure and, at the same time, producing fewer errors.

Big Data and Migration Statistics -

- Inferring Migrations: Traditional Methods and New Approaches based on Mobile Phone, Social Media, and other Big Data (Hughes et al., 2016).
  - LinkedIn: to estimate professional migrations or other labour market indicators for a specific segment of the workforce
    - Not appropriate to obtain information on low-skilled labour migration

Big Data and Migration Statistics – Satellite Data

1.	Departure points
2.	Through traces
3.	External Border monitoring
4.	Coast to the open sea
5.	EU Borders
6.	Conflicts
7.	Internal Movements

7. Mobile Statistics

5. Border Monitoring

6. Conflicts Monitoring

4. Coastal Monitoring & Illegal Vessel

3. Smuggling Hubs Activity Monitoring

2. Tracks Detection

1. Departure Points & Causes

- ESA Big data for migration study - Big Data applications to boost mitigation preparedness and response to migration feasibility study
  - Satellite imagery to identify internal displacement or infrastructural damages after natural disasters (Quinn et.al., 2018).

Big Data and Migration Statistics – Mobile Phone Data

▪ Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda (Blumenstock, 2012)

- ▪ Rwanda's primary telecom operator an exhaustive log of all phone-based activity that occurred from the beginning of 2005 through the end of 2008 (1.5 million subscribers)
  - ▪ time-stamped record of every call that the individual made or received
  - ▪ the closest tower to the subscriber at the time of the transaction (Such as a phone call or text-message)

	All	Farmer	Teacher	Student	Unemployed	Transport
ROG (km)	15.32	13.57*	15.81	18.94*	17.15	26.01**
# towers	14.02	8.99***	11.39**	15.97	15.06	40.04***
Max distance (km)	71.62	60.79***	77.03	74.76	76.69	111.84***
3-month migration	0.23	0.14***	0.28	0.22	0.28	0.39
12-month migration	0.016	0.01	0.01	0.00***	0.02	0.09
N	901	269	109	77	47	23
*Indicates mean of occupation is different from mean of group with $p < 0.05$.						
**Indicates mean of occupation is different from mean of group $p < 0.01$.						
***Indicates mean of occupation is different from mean of group $p < 0.001$.						

Big Data and Migration Statistics – Mobile Phone Data
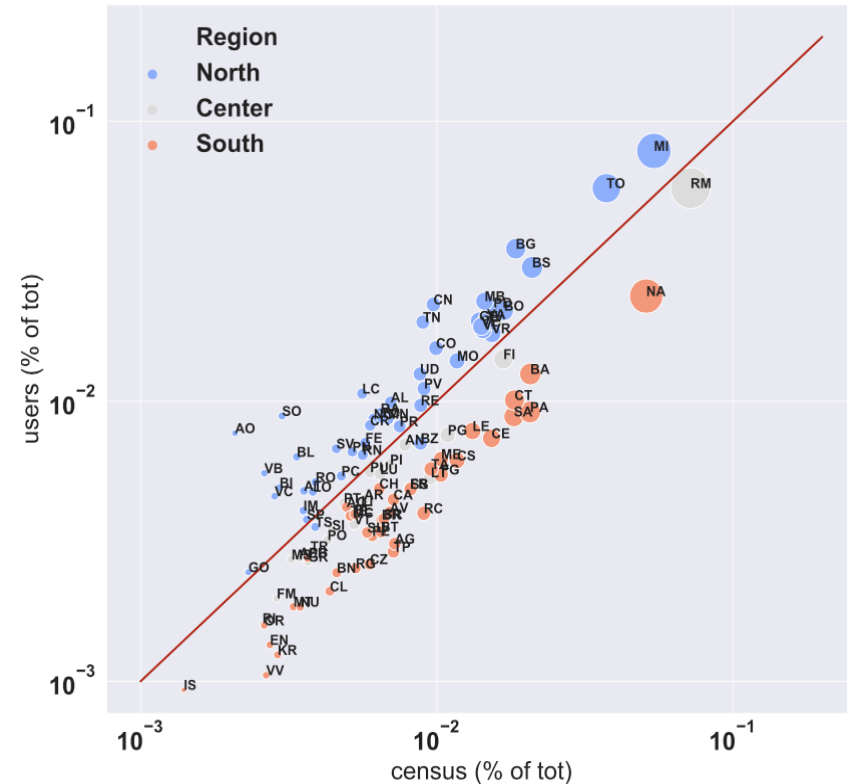
- COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown (Pepe et al., 2020)

- a large-scale dataset of anonymously shared positions of about 170,000 de-identified smartphone users before and during the outbreak, at the sub-national scale

- daily time-series of three different aggregated mobility metrics: the origin-destination movements between Italian provinces, the radius of gyration, and the average degree of a spatial proximity network.
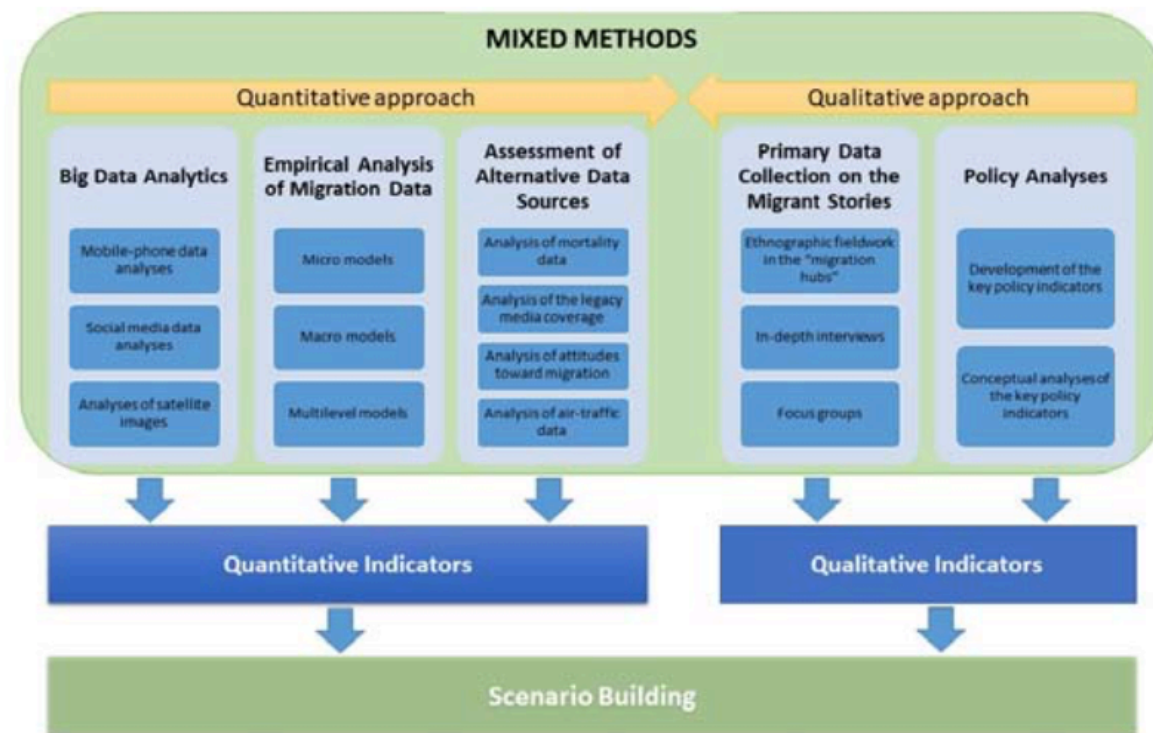
Assessment of Big Data and new methods to study Migration: HumMingBird

▪ HumMingBird Project: Enhanced Migration Measures from a Multidimensional Perspective

  ▪ HumMingBird is a Horizon 2020 project that aims at responding to these needs by improving understandings of changing nature of migration flows and the drivers of migration, by analysing patterns, motivations and new geographies, forecasting emerging and future trends.

Assessment of Big Data and new methods to study Migration: HumMingBird

- 10 countries
  - BE, CH, DE, ES, IT, NL, NO, SE, TR, UK
- 16 partners
  - Research parties/universities
    - Sociology, anthropology, communication science, political science, data science, statistics, telecom engineering, computer sciences
  - SMEs, private companies
  - NGO networks
  - a European Research Infrastructure Consortium (ERIC)

Assessment of Big Data and new methods to study Migration: HumMingBird

- **Migrant stocks:**
  - Twitter and Facebook marketing API data analyses
    - The effect of the social network in attracting immigrants, and how this effect changes from one demographic group to another.
    - Building models similar to the gravity model of mobility, where various other possible drivers are included.
    - An analysis of links to policies and political changes will also be included, e.g. effect of Brexit on highly skilled migration flows.

Assessment of Big Data and new methods to study Migration: HumMingBird

- Seasonal migration:
  - Air-traffic data
    - worldwide air traffic from January 2010 to March 2018 between 239 countries and territories, resulting in a network of 45,444 country pairs.
    - the differences in flows between pairs of countries, paying particular attention to the trend and the seasonality of the movements
    - time-series decomposition that allows to dissect the raw overall air passenger flow between two countries into a trend component and a seasonal component.
  - CDR data analyses

Assessment of Big Data and new methods to study Migration: HumMingBird

- Brain drain/ highly skilled migration
  - LinkedIn and ORCID data analyses, Scientific publications bibliometrics
  - Integrating *data from large professional social networks* such as *LinkedIn*, and *publication data* such as *ORCID*
    - to quantify the brain drain phenomenon both within the EU and also between the EU and the rest of the world.
    - the drivers of highly skilled migration for various groups: work sectors, countries, genders, etc.
    - the effect of the social network in attracting highly-qualified immigrants, and how these changes from one demographic group to another.
  - Scientific publications and author affiliations data
    - To estimate the global mobility of academics

Assessment of Big Data and new methods to study Migration: HumMingBird

- Environmentally induced migration
  - Satellite data analyses (Somalian case)
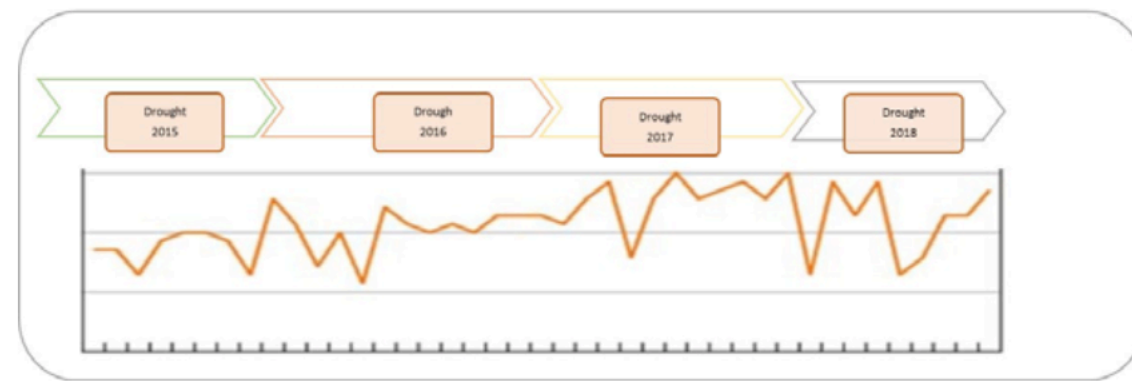   - For the period 2015-2018
    - Sentinel – 1
      - A land use and land cover (LULC) map will be generated at the beginning of each period
        - the changes in land use/ cover due to drought, floods, or violence break outs, will be analysed.
    - Sentinel 2
      - bi-monthly time series will be converted into different time series of Vegetation Indices or indicators

HumMingBird

Assessment of Big Data and new methods to study Migration: HumMingBird

- **Refugees and undocumented migrants**
  - CDR data analyses (Istanbul case)
    - A new dataset in collaboration with Turkey's primary telecommunications operator, Turkcell
      - containing an exhaustive log of all phone-based activity that occurred from the beginning of 2019 through the end of 2019 for a carefully sampled subsample of the population, including refugee and migrant groups
  - Ethics and OPAL
    - The CDR's will be anonymized and aggregated to remove all personal 39 information, along the guidelines established by ethics committees in past endeavours
    - aggregation and anonymization scheme closely follows the Data for Development Challenge setup and the Data for Refugees Challenge setup, where it was shown to provide mobility indicators without jeopardizing privacy or security.

Assessment of Big Data and new methods to study Migration: HumMingBird

- Refugees and undocumented migrants
  - CDR and Nightlights (Satellite) data analyses
    - Antenna traffic data: One-year site-to-site traffic on an 8-hourly basis. This dataset. shows the amount of people from each analysed subgroup (e.g. Syrian or Afghan refugees) at a given location during three intervals of a day.
      - computations and analyses of a number of different metrics related to the migration and mobility patterns of phone owners in Turkey. The living and working locations, as well as social activities for migrants can be determined for the given interval.
    - Nightlights
      - The nightlights are collected and made available by USA's NASA (National Aeronautics and Space Administration) and are available at different time resolutions from annual, monthly and most recently daily.
      - Nightlights dataset for Istanbul will be visualised along with CDR data and other geo-referenced variables (dealing with migration settlements and economic prosperity) in a GIS visualizer.

References

- Bircan, T., Purkayastha, D., Ahmad-Yar, A.W., Lotter, K., Dello Iakono, C., Göler, D., Stanek, M., & Yilmaz, S., Solano, G., & Ünver, Ö. (2020). Gaps in Migration Research. Review of migration theories and the quality and compatibility of migration data on the national and international level (Deliverable n°2.1). Leuven: HumMingBird project 870661 – H2020.
- Blumenstock, J.E., Fafchamps, M., Eagle, N. (2015). Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *J. Dev. Econ.* **120**, 157– 181.
- Böhme, M. H., Gröger, A., Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. Journal of Development Economics, Elsevier, vol. 142(C).
- Hughes, C., Zagheni, E., Abel, G., Wiśniowski, A., Sorichetta, A., Weber, I., Tatem, A.J. (2016). Inferring Migrations: Traditional Methods and New Approaches Based on Mobile Phone, Social Media, and Other Big Data; European Commission: Brussels, Belgium.
- Oliver, N., Letouzé, E., Sterly, H., Delataille, S., De Nadai, M., Lepri, B., Lambiotte, R., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., Salah, A. A., … Vinck, P. (2020). Mobile phone data and COVID-19: Missing an opportunity? ArXiv:2003.12347 [Cs]. https://arxiv.org/ftp/arxiv/papers/2003/2003.12347.pdf
- Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., Tizzoni, M. (2020). COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown. medRxiv. doi: 10.1101/2020.03.22.20039933.
- Quinn, J.A., Nyhan, M.N., Navarro, C., Coluccia, D., Bromley, L. and Oroz. M.L. (2018). Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128):20170363.
- Ruktanonchai, N.W. et al. (2020). Assessing the impact of coordinated COVID-19 exit strategies across Europe. Science.
- Sîrbu, A., Andrienko, G., Andrienko, N. et al. Human migration: the big data perspective. Int J Data Sci Anal (2020). doi: 10.1007/s41060-020-00213-5
- State, B., Rodriguez, M., Helbing, D., Zagheni, E. (2014). Migration of professionals to the U.S.— evidence from LinkedIn data, Social Informatics, 531-274.
- Vinceti, M. et al. (2020). Lockdown timing and efficacy in controlling COVID-19 using mobile 257 phone tracking. EClinicalMedicine. doi:10.1016/j.eclinm.2020.100457
- Zagheni, E., Weber, I. (2012). You are where you e-mail: using e-mail data to estimate international migration rates. In WebSci, pages 348-351.
- Zagheni, E., V.R.K. Garimella, I. Weber, & B. State. (2014). "Inferring international and internal migration patterns from twitter data." In Proceedings of the companion publication of the 23rd international conference on World wide web companion, 439–444. International World Wide Web Conferences Steering Committee.
- Zagheni, E., Weber, I., Gummadi, K., et al. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. Population and Development Review, 43(4):721–734.

Further readings

- Taylor, L., & Meissner, F. (2019). A crisis of opportunity: Market-making, big data, and the consolidation of migration as risk. Antipode, 52(1), 270–290.
- Ruktanonchai, N.W. et al. (2020). Assessing the impact of coordinated COVID-19 exit strategies across Europe. Science.
- Salah, A.A., Pentland, A., Lepri, B. & Letouz´e, E. (2019) .Guide to Mobile Data Analytics in Refugee Scenarios. Springer.
- Sîrbu, A., Andrienko, G., Andrienko, N. *et al.* (2020). Human migration: the big data perspective. *Int J Data Sci Anal.*
- Zagheni, E., Weber, I. (2012). You are where you e-mail: using e-mail data to estimate international migration rates. In WebSci, pages 348-351.

Thanks!

Tuba.Bircan@vub.be

www.hummingbird-H2020.eu
@H20_HumMingBird

HumMingBird

VUB | INTERFACE DEMOGRAPHY RESEARCH GROUP

COORDINATOR

HIVA - Research Institute for Work and Society, KU Leuven (BE)

PARTNERS

University of Utrecht, Department of Information and Computing Sciences (NL) ● University of Salamanca, Faculty of Social Sciences (ES) ● Otto-Friedrich-University of Bamberg, Geographic Migration and Transition Studies (DE) ● CESSDA ERIC (NO) ● EKKE - National Centre for Social Research (GR) ● IEN - Institute of Economic Sciences, Data Center Serbia for Social Sciences (RS) ● Malmö University, Malmö Institute for Studies of Migration, Diversity and Welfare (SE) ● Turkcell Technology (TR) ● GMV Innovating Solutions Ltd, Remote Sensing Services and Explotation Platforms (UK) ● MPG - Migration Policy Group (BE) ● ENoMW - European Network of Migrant Women (BE) ● White Research (BE) ● University of Pisa, Department of Computer Science (IT) ● CNR - National Research Council (IT) ● VUB - Vrije Universiteit Brussel, Interface Demography, Department of Sociology (BE) ● University of Zurich, Department of Social Anthropology and Cultural Studies (CH) ● CILD - Italian Coalition for Civil Liberties and Rights (IT) ● Caritas International (BE)