This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 870661

Use of non-traditional data sources to nowcast migration trends through Artificial Intelligence technologies.

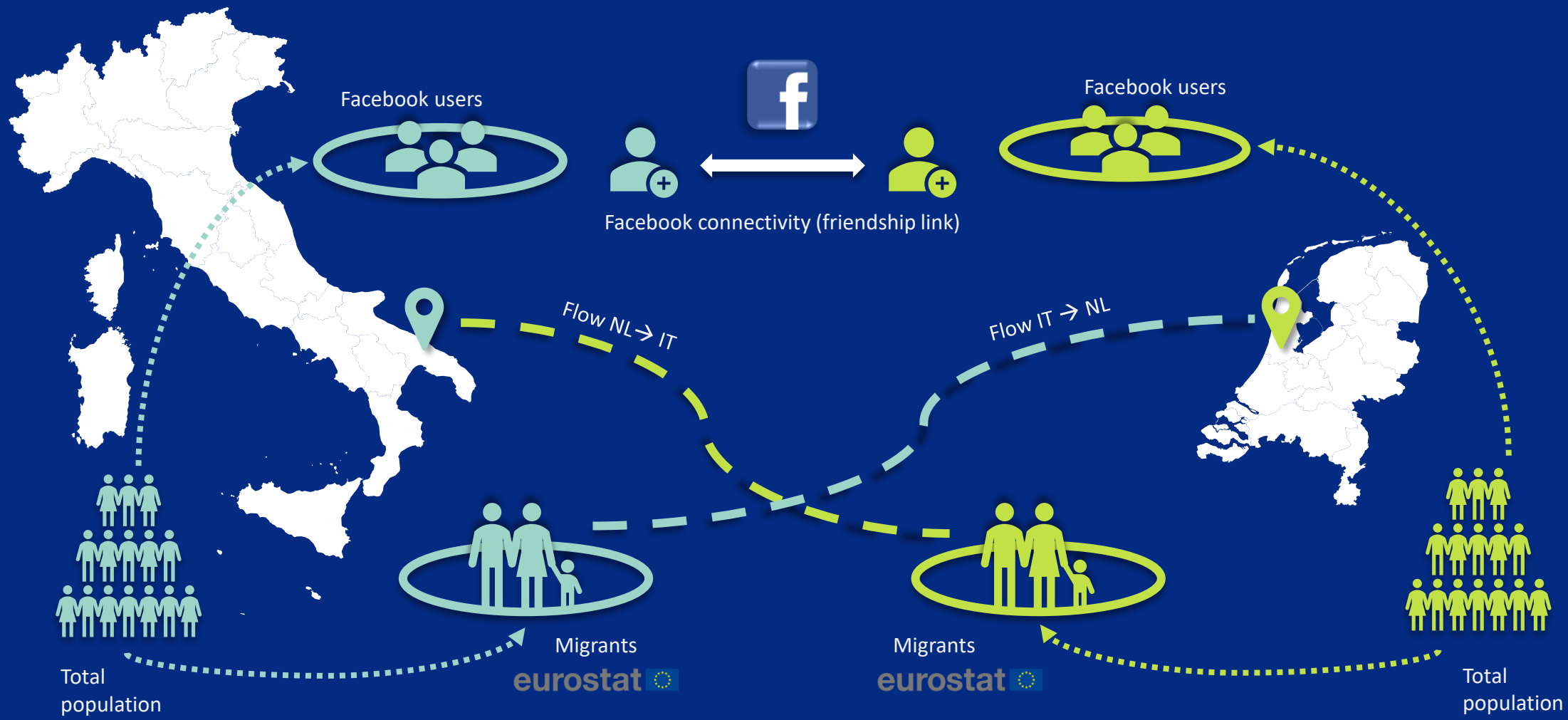
Diletta Goglia, Laura Pollacci, Alina Sîrbu
Computer Science department, University of Pisa, Italy

Session 1a: "How do we measure migration? Methods and advancements".
"Measuring Migration and Mobility" Conference | June 9th and 10th, 2022 |
Nuffield College, University of Oxford, Oxford, GB.

Nowcasting flows with Facebook Social Connectedness

Italy

The Netherlands



Step 1: data

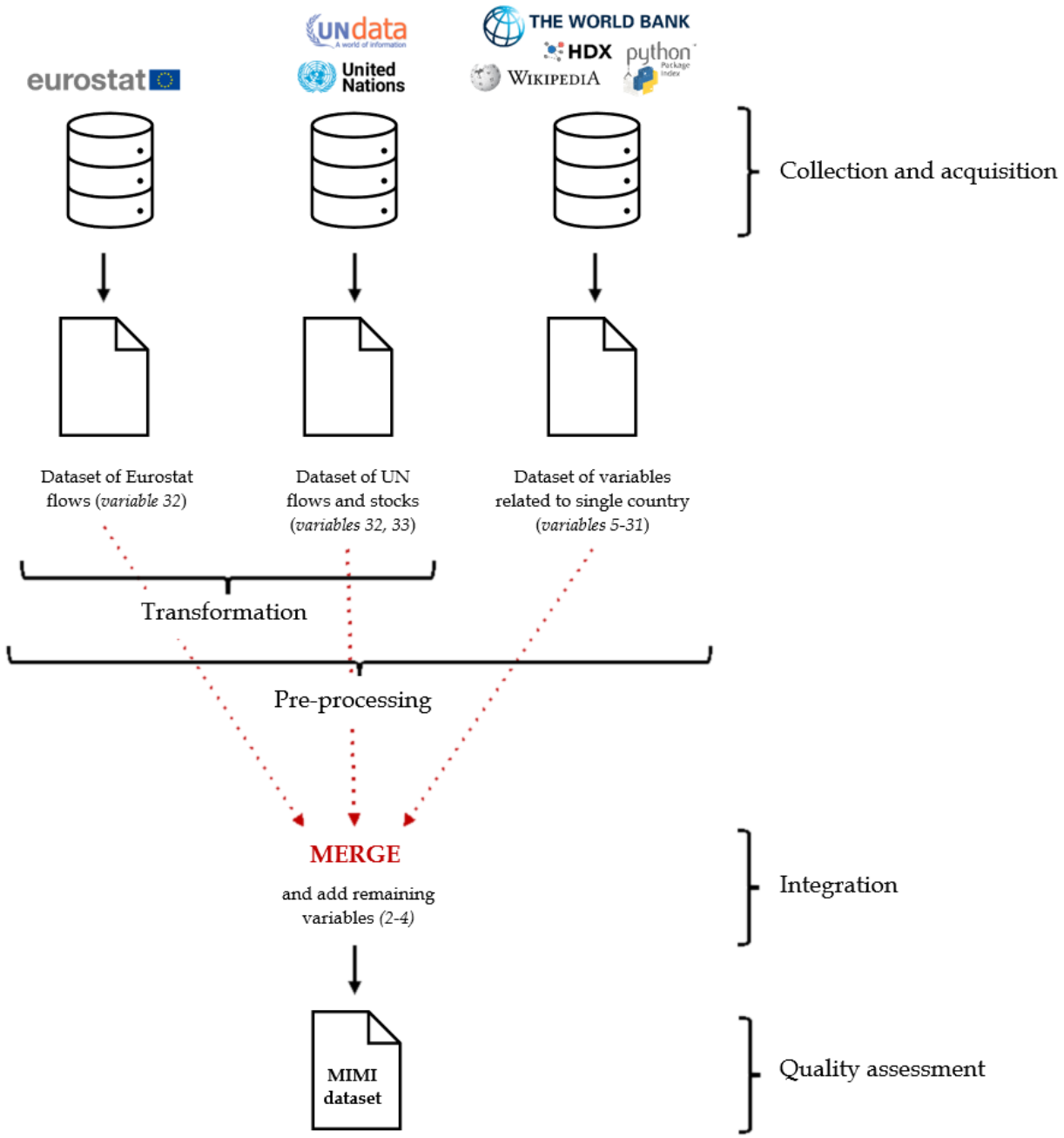
Starting point: MIMI dataset

Multi-aspect Integrated Migration Indicators dataset (CC BY 4.0). Publicly available at [10.5281/zenodo.6493325](https://doi.org/10.5281/zenodo.6493325)

- Non-trivial integration of scattered and heterogeneous data.
- Multidisciplinary variables and original indicators (economic, demographic, cultural and geographic), including Social Network Data.
- Demographers, sociologists, economists and researchers could explore and investigate trends.
- Possibility to develop complex models based on MIMI data.

Related papers

- Goglia, D. (2022) "Multi-aspect Integrated Migration Indicators (MIMI) dataset v2.0", Zenodo, [10.5281/zenodo.6493325](https://doi.org/10.5281/zenodo.6493325).
- Goglia, D., Pollacci, L., Sirbu, A. (2022) "Dataset of Multi-aspect Integrated Migration Indicators", ArXiv pre-print, https://arxiv.org/abs/2204.14223



Step 2: indices

Facebook connectivity & cross-border human mobility

Hypothesis: Facebook strength of connectivity between two countries is related to the bilateral contribution of migration flows.
We built a new indicator (BMI) to nowcast migration flows based on Facebook SCI.

Facebook Social Connectedness Index (SCI)

$$SCI_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

- **Non-traditional** variable within the context of migration studies, included in MIMI for 2020 and 2021 connectivity.
- Provided by "Data for Good at Meta (previously Facebook)" on Humanitarian Data Exchange.
- Measures the **relative probability** that two individuals across two locations i and j are friends with each other on Facebook.
- Could be employed as a proxy of social connections across borders, to be studied as a **possible driver of migration**.
- **Symmetric** structure by definition of "friendship" concept.

Bidirectional Migration Index (BMI)

$$BMI_{i,j} = \frac{Flow_{i\to j}(t) + Flow_{j\to i}(t)}{Pop_i(t) * Pop_j(t)}$$

- i and j represent countries (NUT0 level).
- t is the **reference year**.
- Sum of **direct** and **inverse** contributions → **symmetric** function considering the sharing of both flows.
- Strong **correlation** with SCI for flows provided by Eurostat statistics.
- Can be computed also by sex and by age group.

For couples of countries with inverse flow not available: **Unidirectional Migration Index** (asymmetric – directed – version of BMI).

$$UMI_{i,j} = \frac{Flow_{i,j}}{Pop_i * Pop_j}$$

HumMingBird

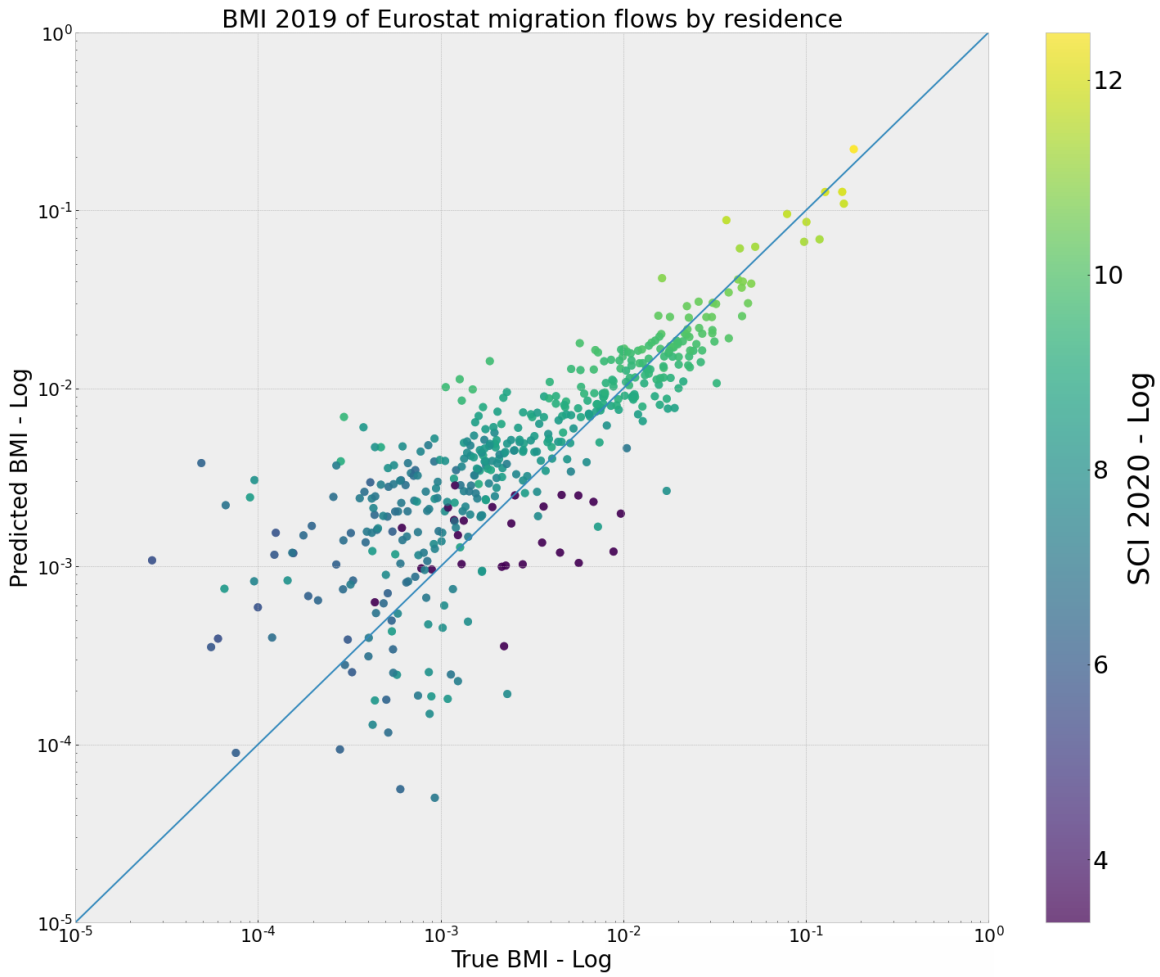
Step 3: fit

Linear Regression

Analysis of migration drivers

Model: Linear regression (Ordinary Least Squares model).
Task: simple fit to determine which variables are relevant and to prove the importance of SCI on BMI prediction.
Aim: feature selection for further neural models.

BMI 2019 of EUROSTAT migration flows by residence: true value VS linear regression predictions (best fit, model with $R^2 = 0.88$). Each data point in the plot represents a couple of countries.



Result

- Facebook strength of connectivity between two countries is strongly positively related to the amount of migration flows they share.
- Higher connectivity, more accurate migration prediction.

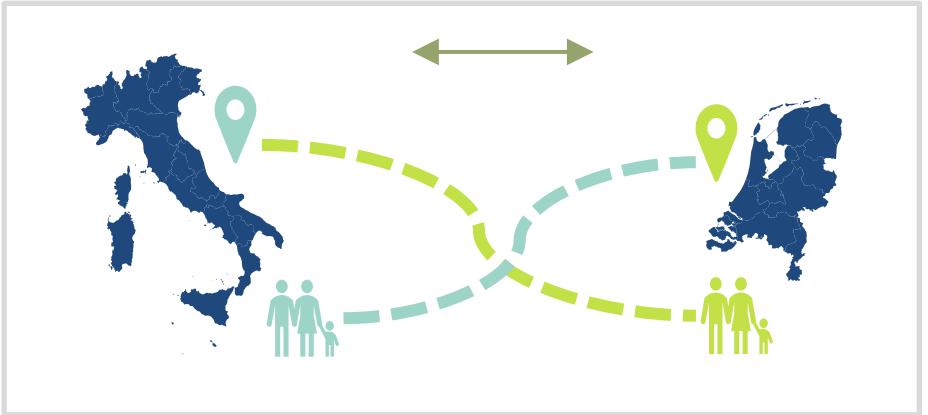
HumMingBird

Results of OLS Backward Elimination Stepwise Linear Regression (I)

Bidirectional Migration Index

	By citizenship (148 couples, 14 countries)				By residence (1114 couples, 63 countries)			
Y	ESTAT BMI 2019 cit, without sci 2020		ESTAT BMI 2019 cit, with sci 2020		ESTAT BMI 2019 res, without sci 2020		ESTAT BMI 2019 res, with sci 2020	
Model n.	1	11	1	6	1	10	1	11
Feature	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|
Intercept	-0.0005		0.0409 ***	0.0412 ***	0.0187 ***	0.0210 ***	0.1079 ***	0.1075 ***
sci_2020			0.0438 ***	0.0438 ***			0.1090 ***	0.1092 ***
geodesic_distance_km	-0.0161 **	-0.0114 ***	-0.0023	-0.0021	-0.0025		9.253e-05	
gdp_diff_2018	0.0074 **	0.0064 **	0.0032 ***	0.0033 ***	0.0040 ***	0.0039 ***	0.0018 ***	0.0016 ***
gdp_mean_2018	-0.0091 *	-0.0074 **	-0.0117 ***	-0.0116 ***	0.0058 ***	0.0055 ***	0.0005	
neighbours	0.0075 **	0.0081 ***	-0.0021 *	-0.0020 **	0.0138 ***	0.0140 ***	-0.0011	-0.0012 *
share_rel	0.0078 ***	0.0070 ***	0.0014	0.0013	0.0006		-0.0005 *	-0.0005 *
share_lang	-0.0024		-0.0013 *	-0.0014 **	0.0020 ***	0.0020 ***	0.0004	0.0004
PDI_diff	-0.0038	-0.0034	0.0020	0.0020	-0.0025 *	-0.0026 *	-0.0006	
IDV_diff	-0.0067 **	-0.0065 **	-0.0021 *	-0.0021 *	-0.0012		-0.0017 ***	-0.0017 ***
UAI_diff	0.0014		-0.0033 *	-0.0033 *	0.0003		0.0006	
MAS_diff	-0.0103 ***	-0.0088 ***	-0.0021 *	-0.0021 *	-0.0051 ***	-0.0050 ***	-0.0017 ***	-0.0018 ***
fb_users_diff	-0.0018		-0.0023 *	-0.0024 ***	0.0066	0.0068	-0.0007	
fb_users_perc_diff	0.0025		0.0043 ***	0.0044 ***	-0.0006		-8.071e-05	
fb_users_perc_mean	0.0028	0.0049	0.0025	0.0026 *	0.0054 ***	0.0055 ***	0.0044 ***	0.0047 ***
fb_users_mean	-0.0134 **	-0.0145 ***	-0.0005		-0.0114 *	-0.0112 *	-0.0010	-0.0016 **
area_diff	-0.0019		0.0002		-0.0029		0.0012	
area_mean	0.0036		0.0004		0.0034		-0.0011	
share_cont	0.0005				0.0047 ***	0.0063 ***	0.0010	0.0011 ***
R2 (centered)	0.574	0.560	0.949	**0.949**	0.363	0.361	0.880	**0.880**
AIC	-846.3	-855.7	-1158.	**-1164.**	-6321.	-6332.	-8179.	**-8191.**
BIC	-795.4	-825.7	-1104.	**-1119.**	-6231.	-6276.	-8084.	**-8136.**

Previous plot



- Dependent variable (y) is BMI → Symmetric indicator → Independent variables are undirected.
- Four different models: flows by citizenship / residence, including / not including SCI.
- Substantial improvement in models with SCI.
- SCI positively and strongly relevant feature.

```
  * p-value < 0.05
 ** p-value < 0.01
*** p-value < 0.001
```

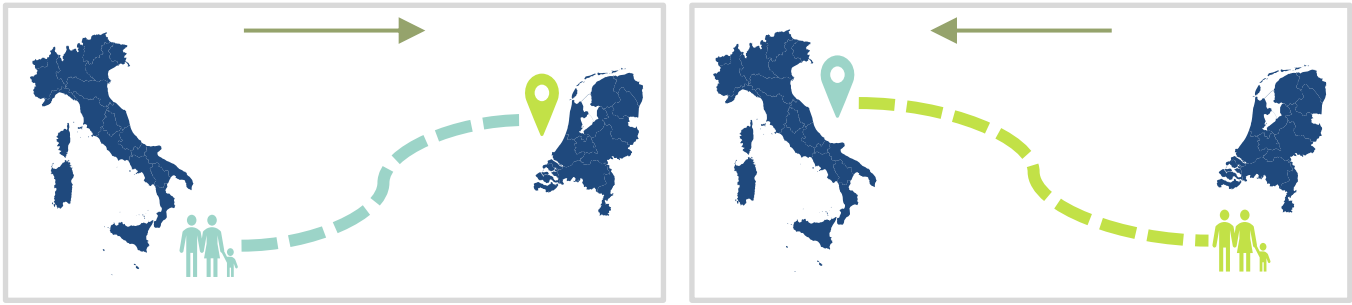
Results of OLS Backward Elimination Stepwise Linear Regression (II)

Unidirectional Migration Index

	By citizenship (748 couples, 200+ countries)			
y	ESTAT UMI 2019 cit, without sci 2020		ESTAT UMI 2019 cit, with sci 2020	
Model n.	1	20	1	18
Feature	coef, P>|t|	coef, P>|t|	coef, P>|t|	coef, P>|t|
intercept	-0.0027		0.2080 ***	0.2072 ***
sci_2020			0.2024 ***	0.2024 ***
origin_PDI	-0.0056		-0.0009	
destination_PDI	0.0012		0.0036*	
origin_IDV	-0.0053		-0.0011	
destination_IDV	-0.0021		-0.0051	
origin_UAI	-0.0033		-0.0017	
destination_UAI	-0.0033		-0.0049 *	-0.0023 *
origin_MAS	-0.0035	-0.0053	-0.0005	
destination_MAS	0.0022		-0.0002	
origin_area	0.0002		0.0004	
destination_area	-0.0017		-0.0040 *	-0.0020 *
origin_fb_users	-0.0012		-0.0012	
destination_fb_users	-0.0038	-0.0071 ***	0.0036	
origin_fb_users_perc	0.0054	0.0043	0.0039 *	0.0025
destination_fb_users_perc	0.0012		0.0010	
geodesic_distance_km	-0.0114 *	-0.0111 ***	-0.0025	-0.0038 **
origin_gdp_2018	0.0007		-0.0015	
destination_gdp_2018	0.0026		0.0026	
gdp_diff_2018	-0.0002		-0.0018	
neighbours	-0.0049		0.0042 **	0.0040 **
share_cont	0.0010		0.0017	
share_rel	0.0031	0.0033 **	0.0009	0.0013 *
share_lang	0.0084 ***	0.0077 ***	0.0010	0.0012
R2 (centered)	0.115	0.103	0.780	**0.777**
AIC	-3043.	-3065.	-4081.	**-4100.**
BIC	-2941.	-3037.	-3975.	**-4059.**

GDP proxy

GDP proxy



- Dependent variable (y) is UMI → Asymmetric indicator → Direction preserved in independent variables (distinction kept between origin and destination).
- Strong improvement in the model with SCI.
- SCI is again a very relevant feature, with positive coefficient.
- Other relevant features are different w.r.t. the previous model.
- GDP does not appear but remains significant since approximated by the presence of «country area» and «FB users» variables.

```
 * p-value < 0.05
** p-value < 0.01
*** p-value < 0.001
```

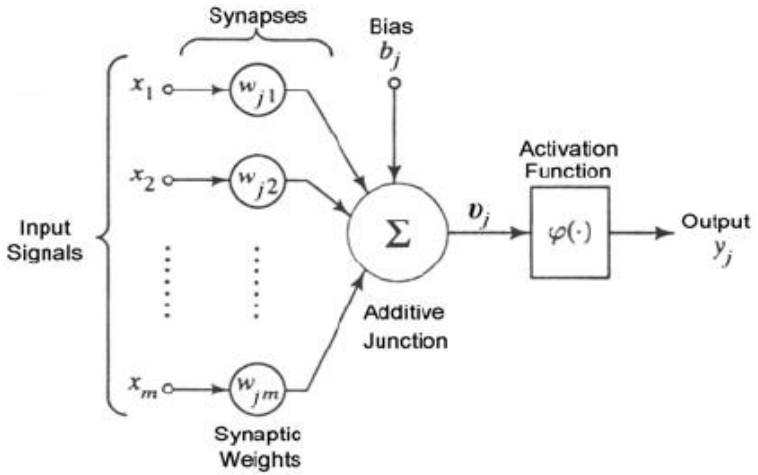
Next step: Deep Learning model

New tasks: migration **nowcasting** and **prediction** with deep Artificial Neural Networks (e.g., Feed Forward Multilayer Perceptron).
Input layer = relevant features selected by regression model + migration time series data.

Two possible analyses:

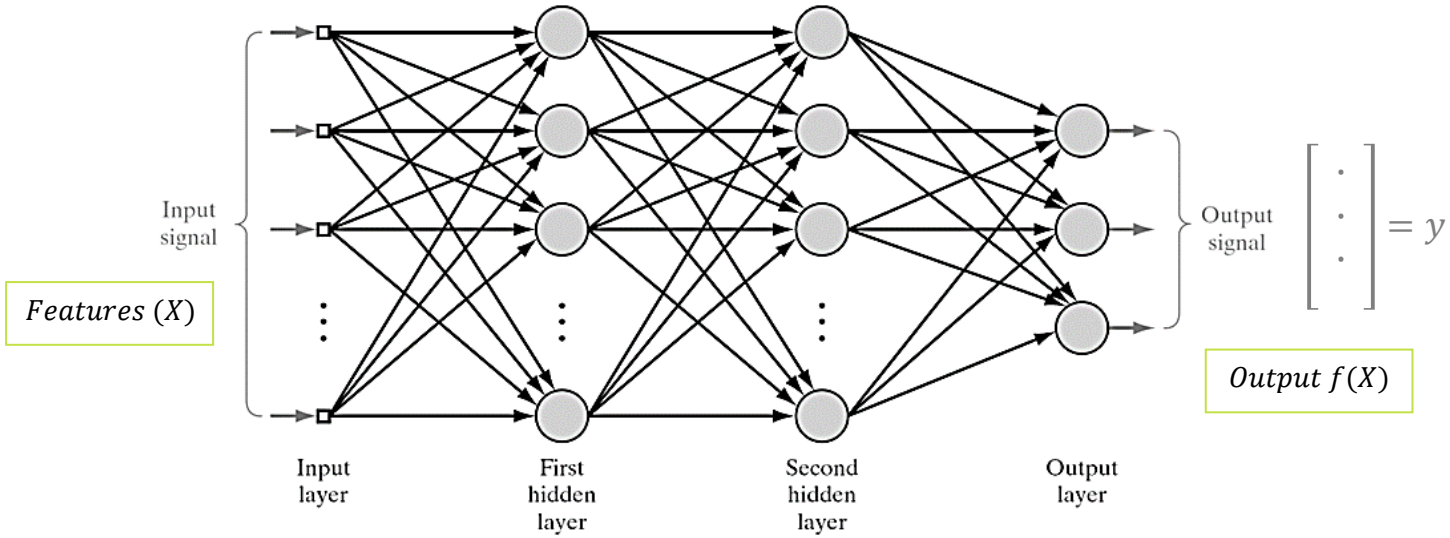
- Exploit a time window of past flows to predict future migration.
- Train-test split on couples of countries to predict migration flows only for some pairs.

Scheduled: experiments on MLP and Random Forest models and evaluation on better performance outcomes.



$$y_j = \varphi(\sum_{k=1}^{m} w_{jk} x_k + b_j)$$

Artificial neuron



Standard structure of MLP

HumMingBird

HumMingBird

AI4MIG

Want more? Scan me!

Contact

Dr. Diletta Goglia
Postgraduate Student in MSc in Artificial Intelligence
Computer Science department, University of Pisa, Italy
d.goglia@studenti.unipi.it
www.dilettagoglia.netlify.app

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 NSL Limited (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)

Horizon2020
European Union Funding
for Research & Innovation

UNIVERSITÀ DI PISA

SoBigData

VISIT: http://www.hummingbird-H2020.eu

CONTACT US: HumMingBird@kuleuven.be

FOLLOW US:

@HumMingBird.H2020
HumMingBird Project
@H20_HumMingBird
HumMingBird Project