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

Inferring migration patterns from social media data

WP5 Leader: UNIPI–CNR(ISTI)

Participants: USAL, UU, VUB

5th Consortium Meeting
9-10 December 2021
Salamanca

T5.1 Proxy for migration stocks using Facebook data

Technical specification about our datasets:

- **Bilateral** migration **flows** 2018-20
- **NUTS0** level (country to country)
- 232 different countries
- Final result: huge matrix of flows for explorative analysis
- 19k+ records × 73 features (World)
- 40k+ records × 77 features (Europe)
- Returners excluded (flows from **/** to the same country)
- Migrants' organization in three disjoint clusters based on EUROSTAT classification system[2]: by country of **citizenship**, **birth** & prev/next **residence**
- Reference year for today qualitative analysis: 2019.

Data Integration

- **Demographic** stats
- Geographical info (standardised to ISO3166)
- Pre-existing migration indicators (NET rate, total volume, ...)
- **New migration indices** built from scratch

Data sources:

Humanitarian Data Exchange FACEBOOK Data for Good	Facebook connectivity (SCI), updated **October 2021**
UNdata A world of information	World migration flows, population stocks, NET
eurostat	European migration flows & population stocks
United Nations Statistics Division	Integrate missing values
THE WORLD BANK IBRD • IDA	Integrate missing values

HumMingBird

Indices definition

& interpretation as relative probabilities

- **Social Connectedness Index (SCI)** between two locations i and j measures the relative probability that two individuals across the two locations are friends with each other on Facebook[1]. It is defined as:

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

- Our **Migration Index (MI)** between two locations i and j measures the relative probability of a person, *permanently* resident [2] in i or j, to be a migrant from location i to j (or vice versa). It is defined as:

$$MI_{i,j} = \frac{(migr\ flow)_{i,j} + (migr\ flow)_{j,i}}{(pop\ stock)_i * (pop\ stock)_j}$$

Methodology

- Some indices built exploiting available indicators: NET migrations estimates and projections, (difference between) migration volumes, … Best correlation values obtained for the final MI (left) which uses values of **flows.**
- Sum of **direct** and **inverse** flows as they both contribute to determine the migration between the two locations.
- SCI is **symmetric** by definition of FB “friendship” so MI has been built to be symmetric too (in reality: inverse flows not always available).
- Both indices are scaled to have a max of 1billion and a min of 1.

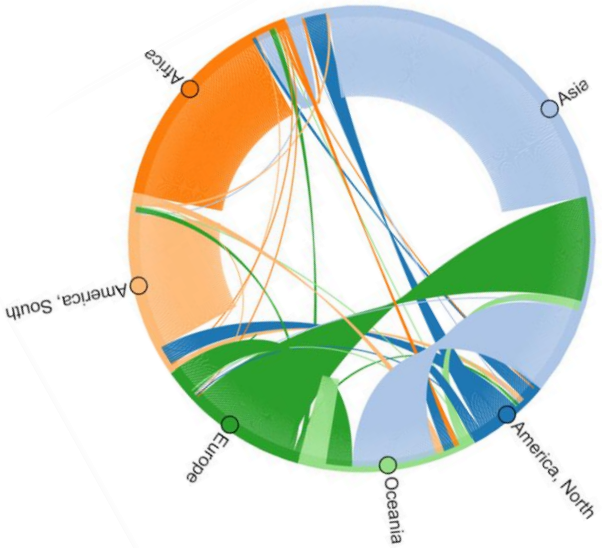
N.B. Later changes: consider including geodesic distance [3] between i and j in the denominator to penalise the strength of connections between neighbouring countries (better correlations noticed). **[discarded]**

HumMingBird

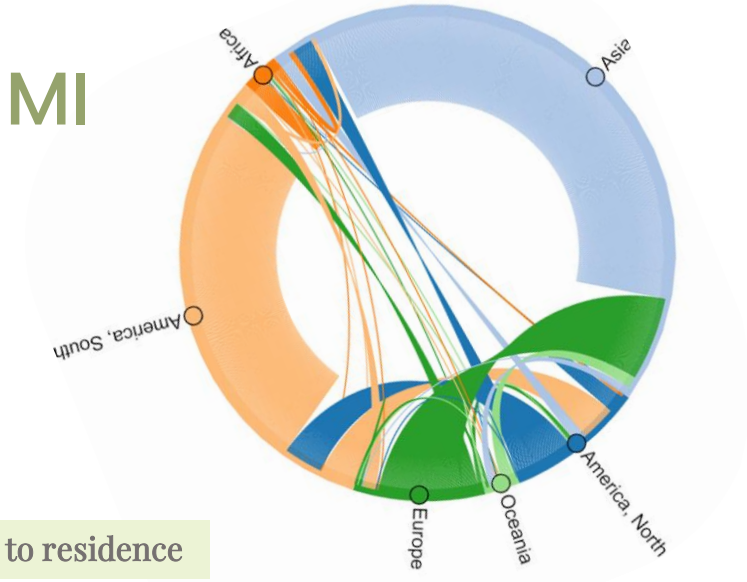
[1] “Documentation - FB Social Connectedness Index - October 2021”, data.humdata.org
[2] “usual resident inhabitants” based on the definition of "total population" contained in Eurostat, Glossary on Demographic Statistics.
[3] computed by using countries’ centroids’ coordinates

World

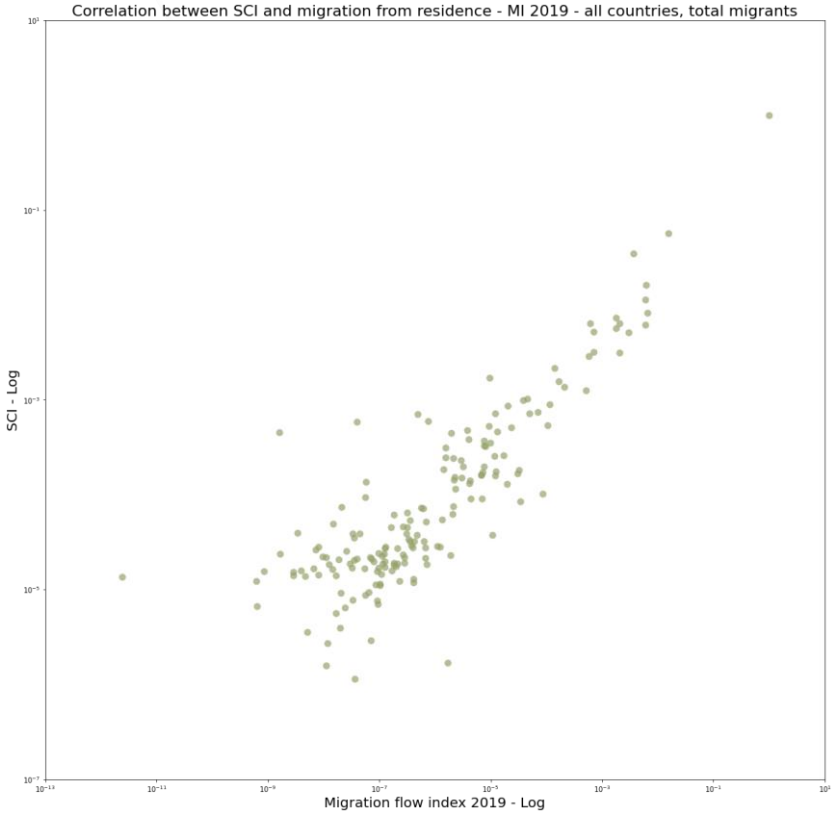
Correlation between SCI and MI



Main focus: find out potential existing correlations between migration patterns and Facebook connectivity.

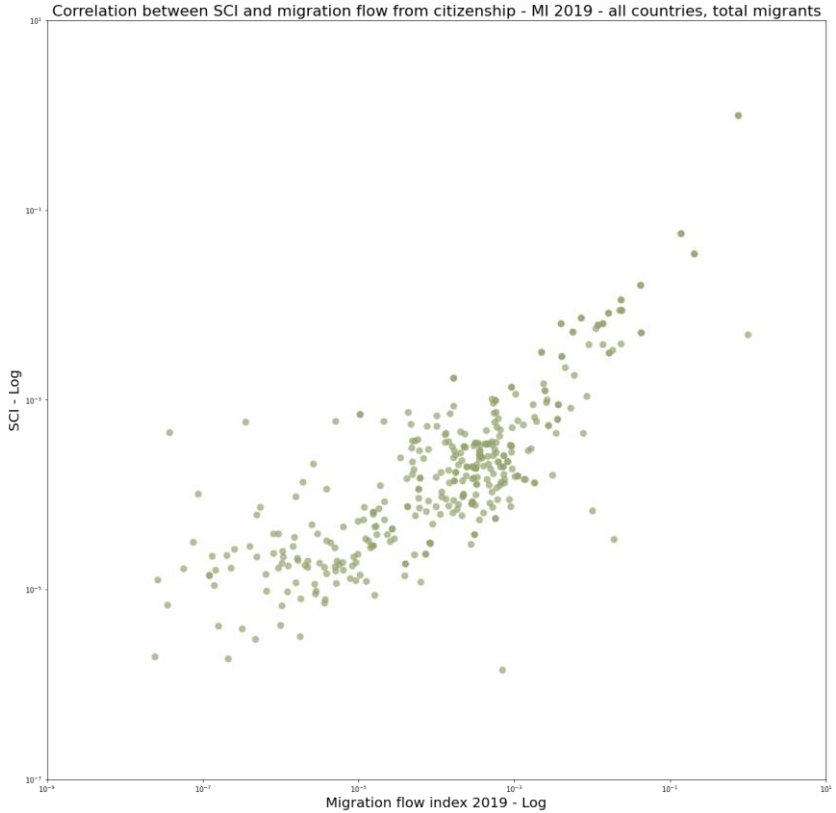


Change of residence



Correlation SCI-MI			
		From	
		Res	Cit
Method	Pearson	0.71	0.43
	Kendall	0.28	0.26
	Spearman	0.44	0.37

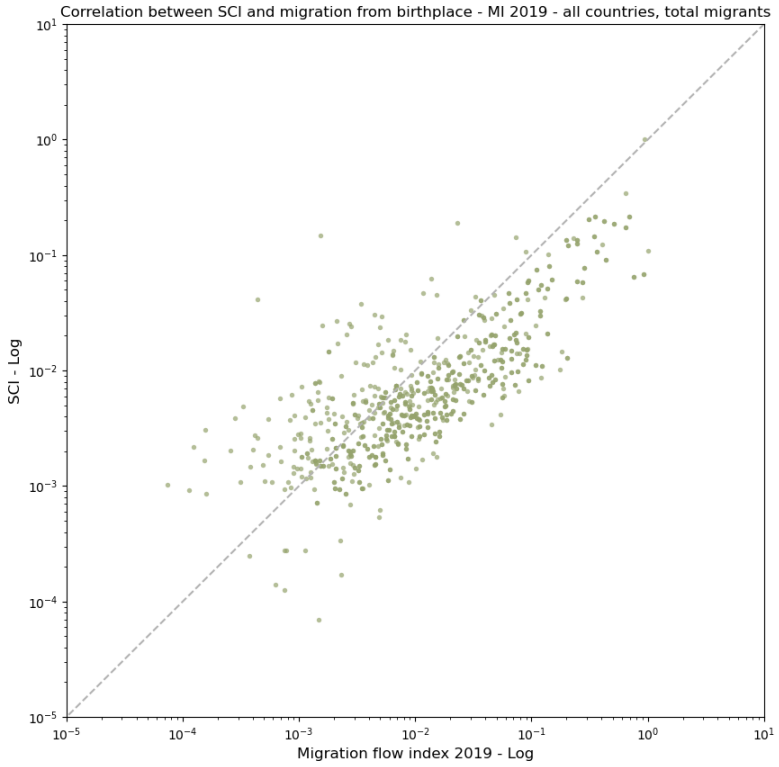
From citizenship to residence



Europe

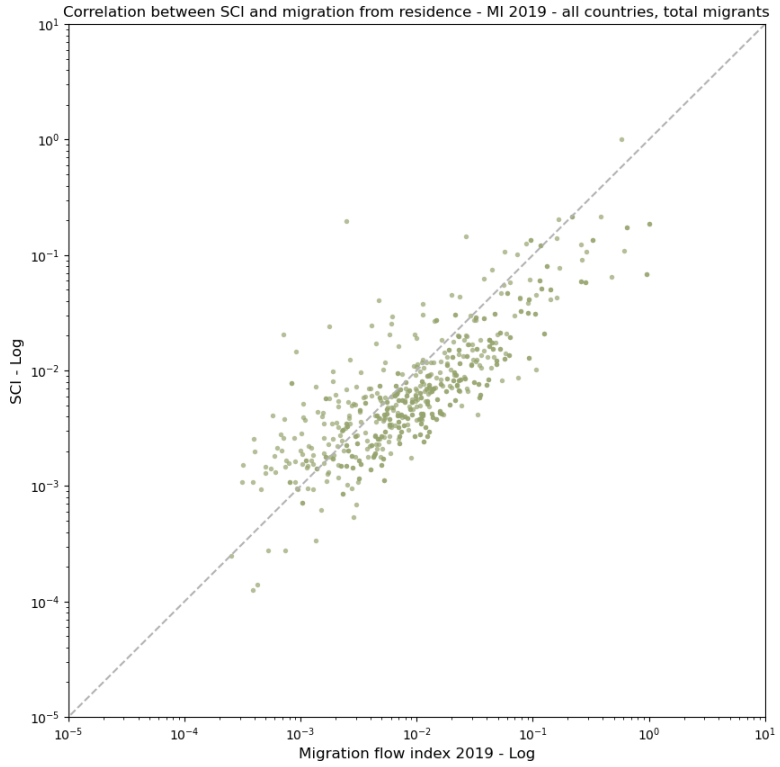
Correlation between SCI and MI

From birthplace



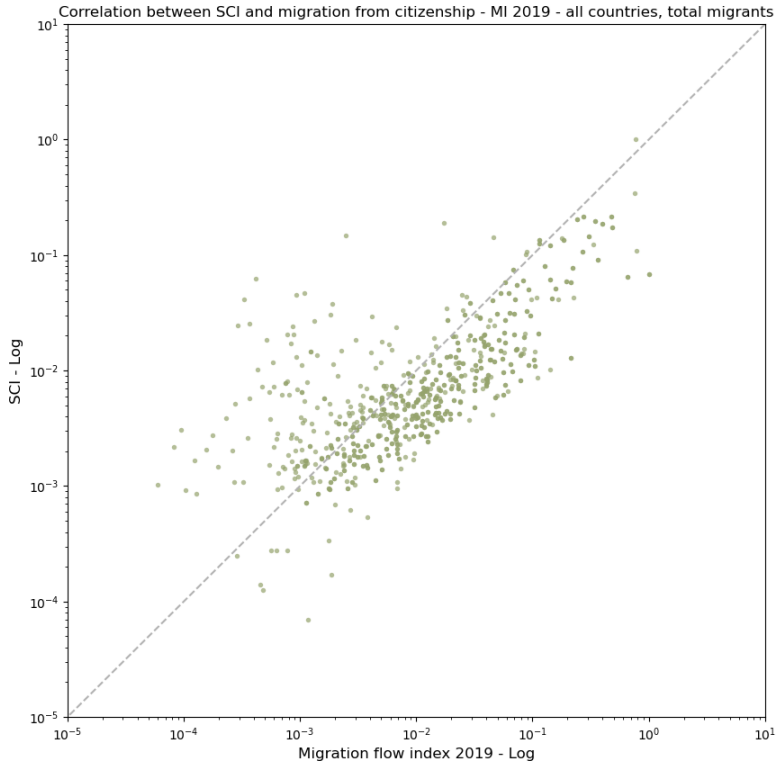
Correlation SCI-MI	
Method	Value
Pearson	0.7
Kendall	0.6
Spearman	0.8

From residence



Correlation SCI-MI	
Method	Value
Pearson	0.6
Kendall	0.6
Spearman	0.8

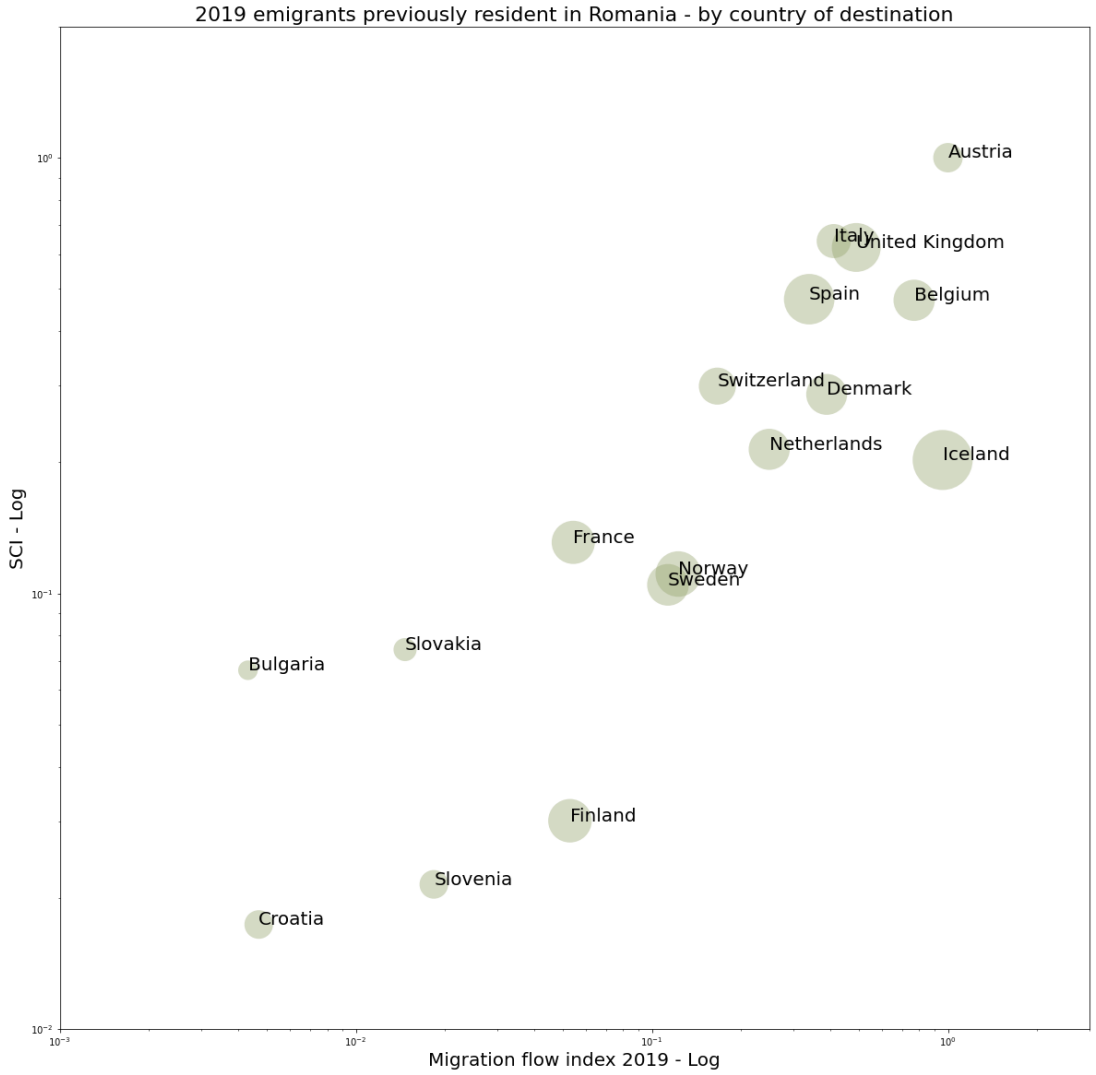
From citizenship

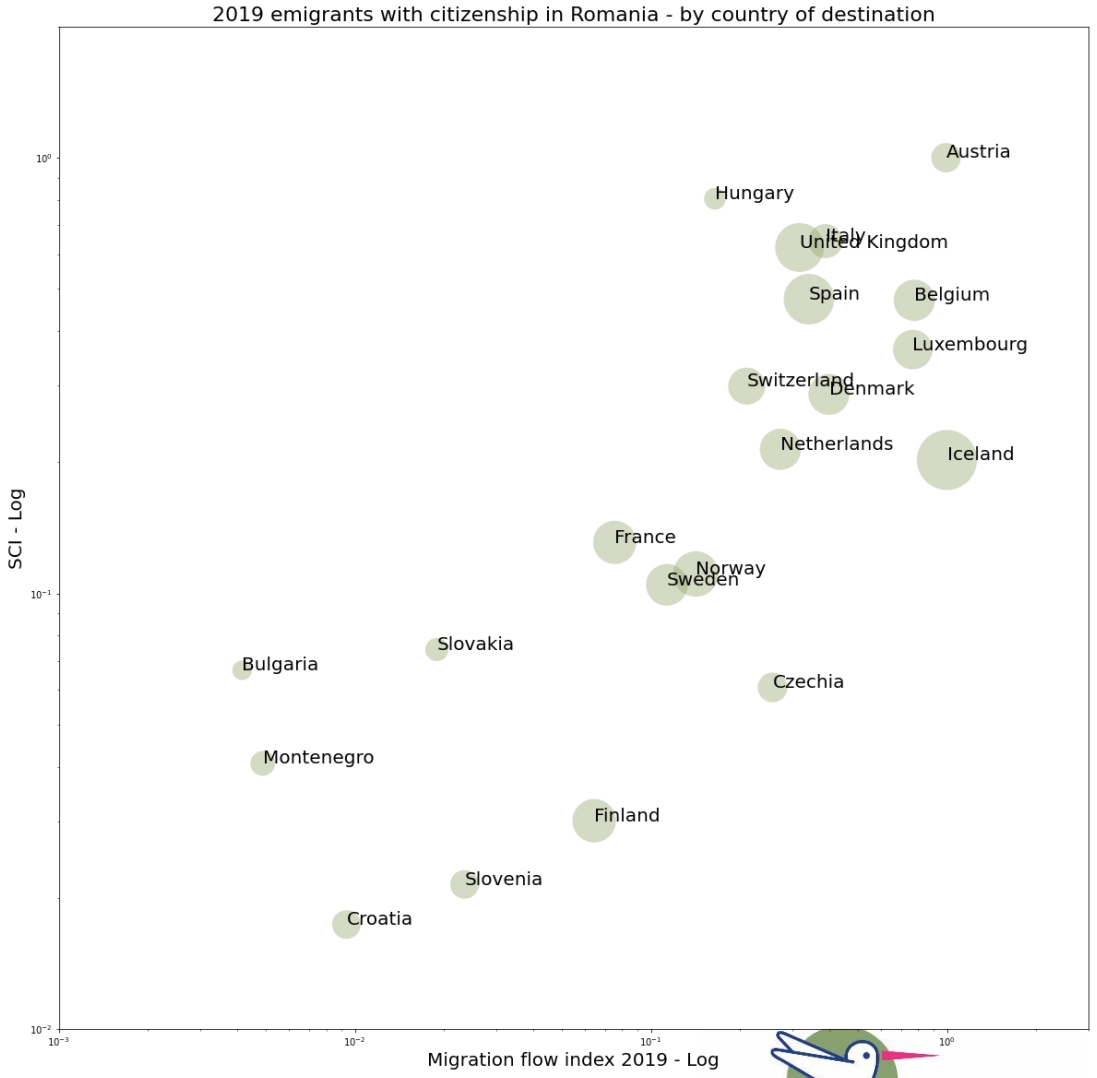


Correlation SCI-MI	
Method	Value
Pearson	0.7
Kendall	0.6
Spearman	0.7

Distance between countries

Romanian migration outflow (and connectivity) in Europe

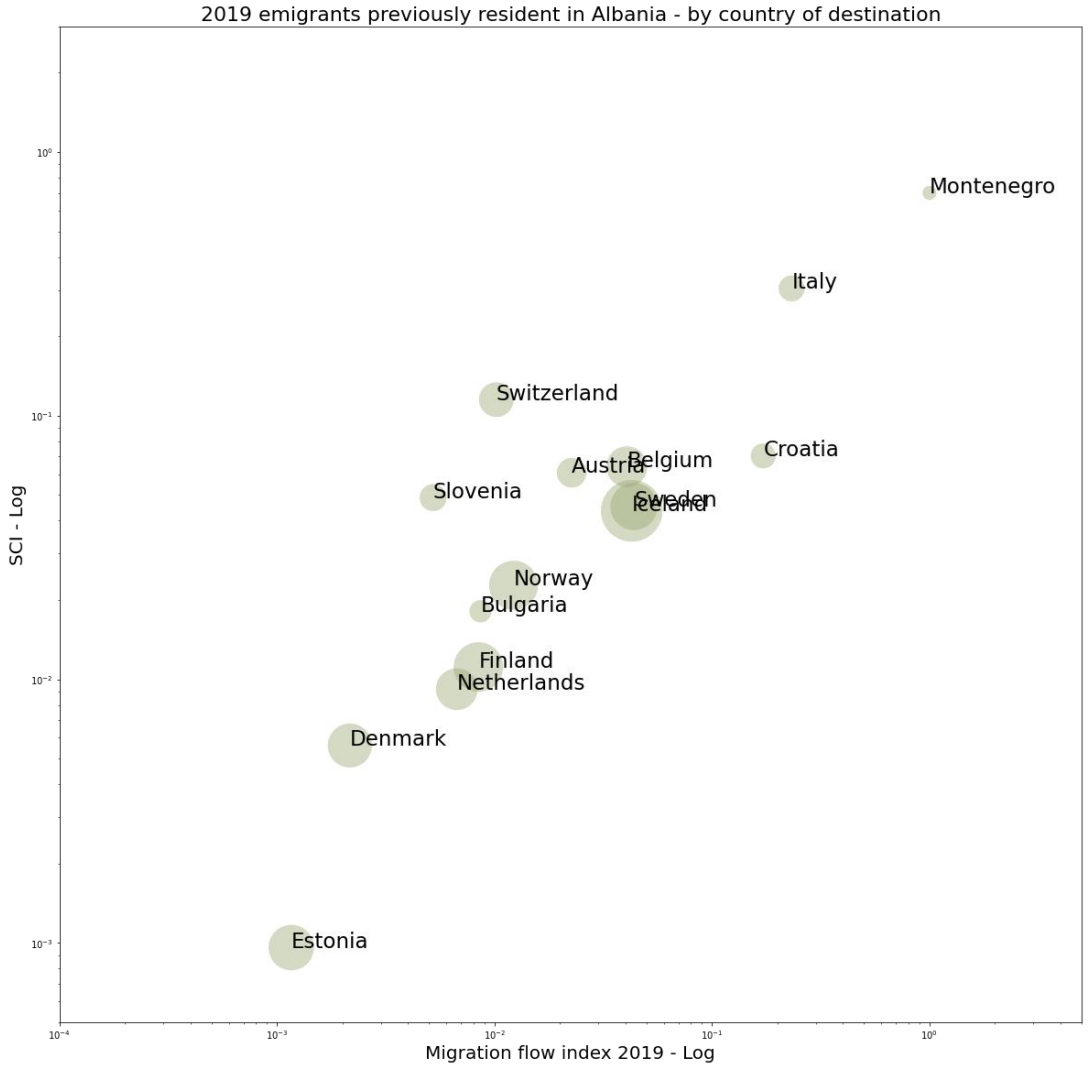


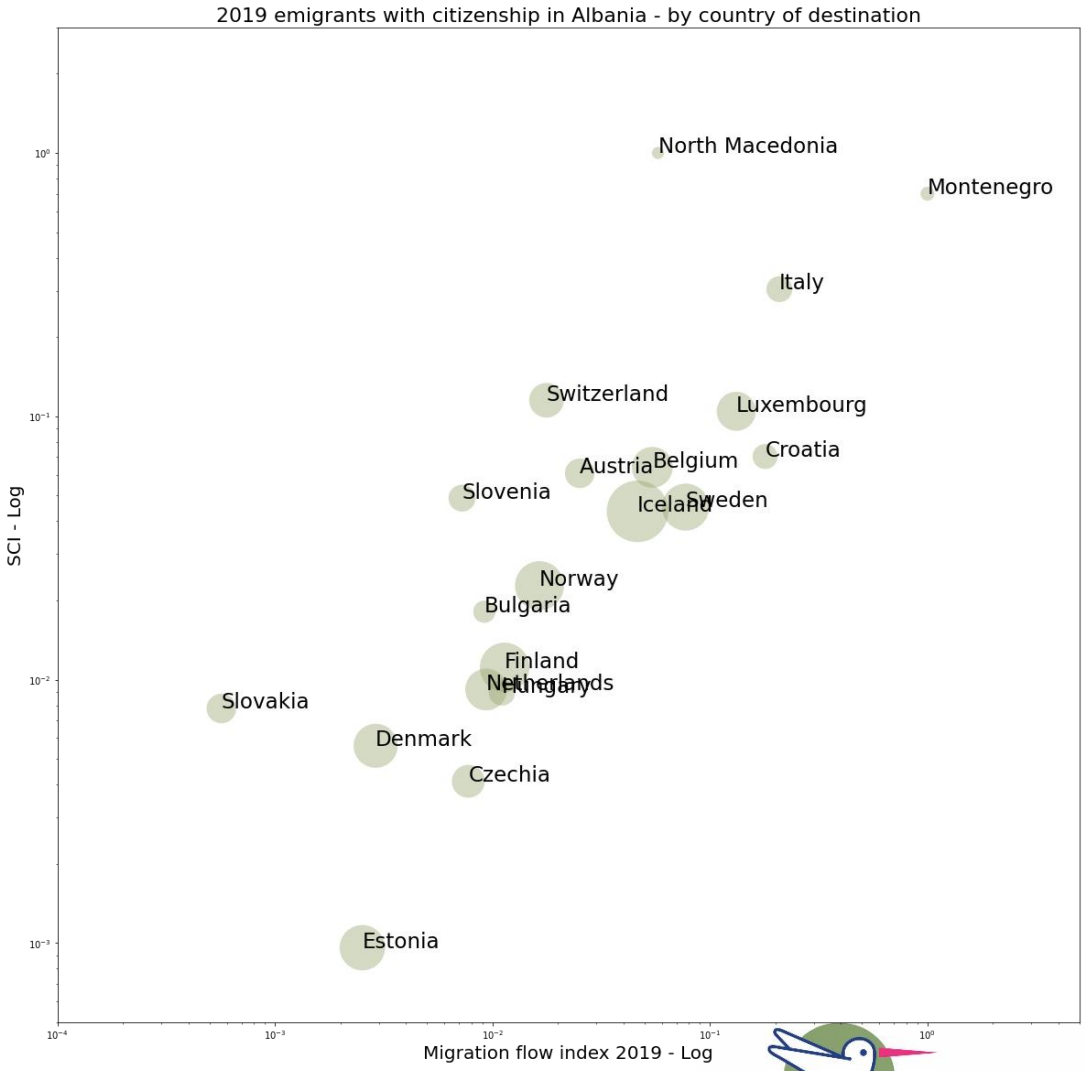


HumMingBird

Distance between countries

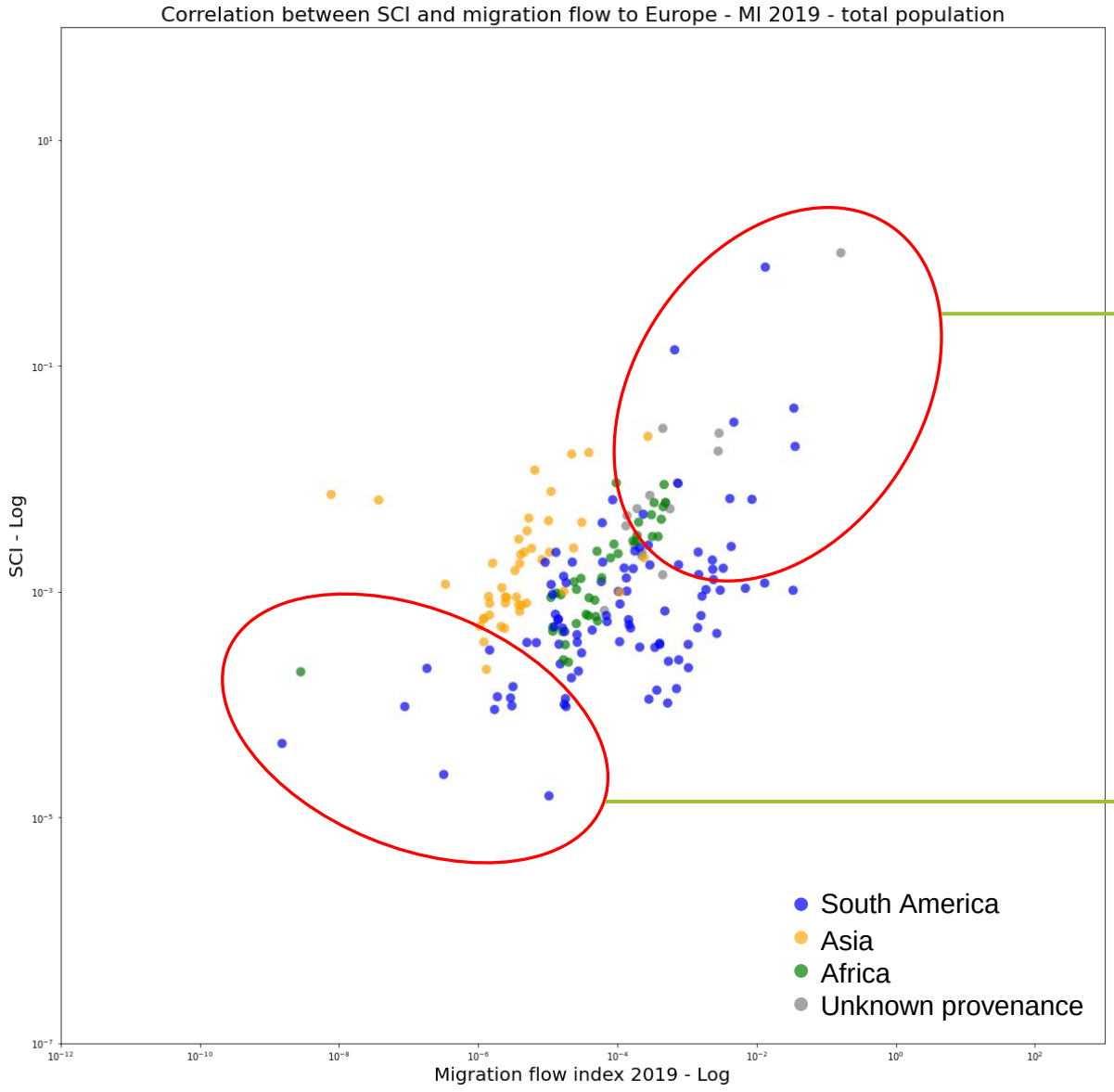
Albanian migration outflow (and connectivity) in Europe





HumMingBird

International citizens moving their residence (and creating FB connections) in Europe in 2019



Destination: West Europe (Belgium, Netherlands, Luxembourg, Germany, Denmark, UK, Switzerland, Austria, Spain, Portugal)

Destination: East Europe (Republic of Moldova, Poland, Ukraine, Estonia, Bosnia and Herzegovina, Slovakia, Hungary)

Correlation SCI-MI	
Method	Value [5]
Pearson	0.8
Kendall[6]	0.3
Spearman[6]	0.4

[5] rounded up to the nearest first decimal
[6] Log value & 'omit' option as NaN policy: performs the calculations between logs, ignoring NaN values

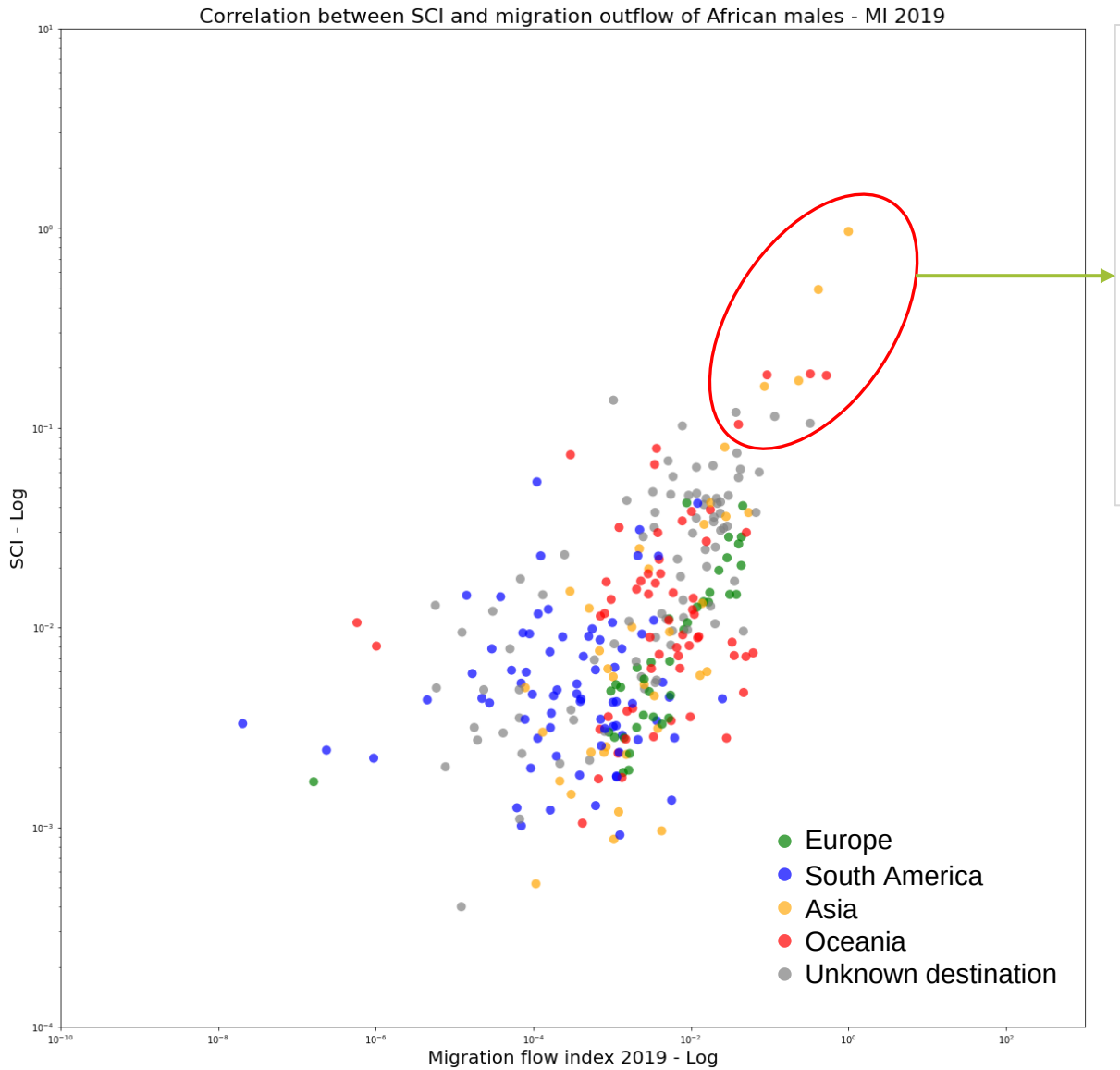
HumMingBird

African male citizens moving their residence (and creating FB connections) worldwide in 2019

	provenance	destination
0	Egypt	Kuwait
1	Kenya	Qatar
2	Sierra Leone	Kuwait
3	Kenya	Bahrain
4	Ethiopia	Kuwait
5	Madagascar	Kuwait
6	Seychelles	Kuwait
7	Mauritius	Australia
8	Seychelles	Australia
9	Seychelles	Fiji
10	Kenya	United Arab Emirates
11	Kenya	Saudi Arabia
12	Benin	Fiji
13	Kenya	Haiti
14	Burundi	Canada
15	Djibouti	Canada
16	Mauritius	Canada
17	Zimbabwe	Australia
18	Kenya	Jamaica
19	Libyan Arab Jamahiriya	Kuwait

Top-20 pairs of countries sorted by highest values of SCI and MI 2019.

Correlation SCI-MI	
Method	Value
Pearson	0.5
Kendall	0.4
Spearman	0.6



Destination: world countries with **high GDP** (gross domestic product) **at PPP (**purchasing power parity) **per capita.**[4]

Mostly **Western Asia** (Qatar, United Arab Emirates, Bahrain, Oman, Australia, Canada) preferred for distance and cultural similarity reasons (i.e. religion).

HumMingBird

[4] i.e. the purchasing power parity (PPP) value of all final goods and services produced within a country in a given year, divided by the average population for the same year.

Colonies

Migration flows to/from Spain

International citizens

	provenance	destination	MI_2019	sci
0	Cuba	Spain	386,983.86824	nan
1	Peru	Spain	15,510.57461	13,932.00000
2	Ecuador	Spain	14,882.53624	30,067.00000
3	Philippines	Spain	10,061.36801	1,568.00000
4	Colombia	Spain	2,045.82491	22,668.00000
5	Aruba	Spain	195.59179	20,032.00000
6	Uzbekistan	Spain	191.42113	508.00000
7	Åland Islands	Spain	122.17209	nan
8	Faroe Islands	Spain	100.99846	nan
9	Kenya	Spain	21.53141	858.00000

	provenance	destination	MI_2019	sci
0	Spain	Viet Nam	1,247,096.14941	242.00000
1	Spain	Kuwait	315,532.25560	1,636.00000
2	Spain	Argentina	23,066.44144	16,376.00000
3	Spain	Peru	15,510.57461	13,932.00000
4	Spain	Ecuador	14,882.53624	30,067.00000
5	Spain	Thailand	3,455.15075	375.00000
6	Spain	Colombia	2,045.82491	22,668.00000
7	Spain	Guernsey	398.60727	nan
8	Spain	Aruba	195.59179	20,032.00000
9	Spain	Canada	155.02321	3,070.00000

European citizens

	provenance	destination	MI_2019	sci
0	France	Spain	9.115723e+06	10218.0
1	Italy	Spain	1.607655e+07	17626.0
2	United Kingdom	Spain	1.194262e+07	13252.0
3	Romania	Spain	3.771340e+07	50921.0
4	Belgium	Spain	2.093445e+07	14593.0
5	Portugal	Spain	2.296671e+07	29713.0
6	Germany	Spain	3.078374e+06	9782.0
7	Poland	Spain	2.502049e+06	3559.0
8	Netherlands	Spain	1.563316e+07	11482.0
9	Bulgaria	Spain	2.175125e+07	21141.0

	provenance	destination	MI_2019	sci
0	Spain	France	9.115723e+06	10218.0
1	Spain	Romania	2.978563e+07	50921.0
2	Spain	Italy	1.607655e+07	17626.0
3	Spain	Netherlands	1.563316e+07	11482.0
4	Spain	Hungary	8.959139e+05	4345.0
5	Spain	Czechia	9.621242e+05	3800.0
6	Spain	Switzerland	1.881887e+07	23340.0
7	Spain	Bulgaria	2.175125e+07	21141.0
8	Spain	Denmark	9.435471e+06	7521.0
9	Spain	Belgium	2.093445e+07	14593.0

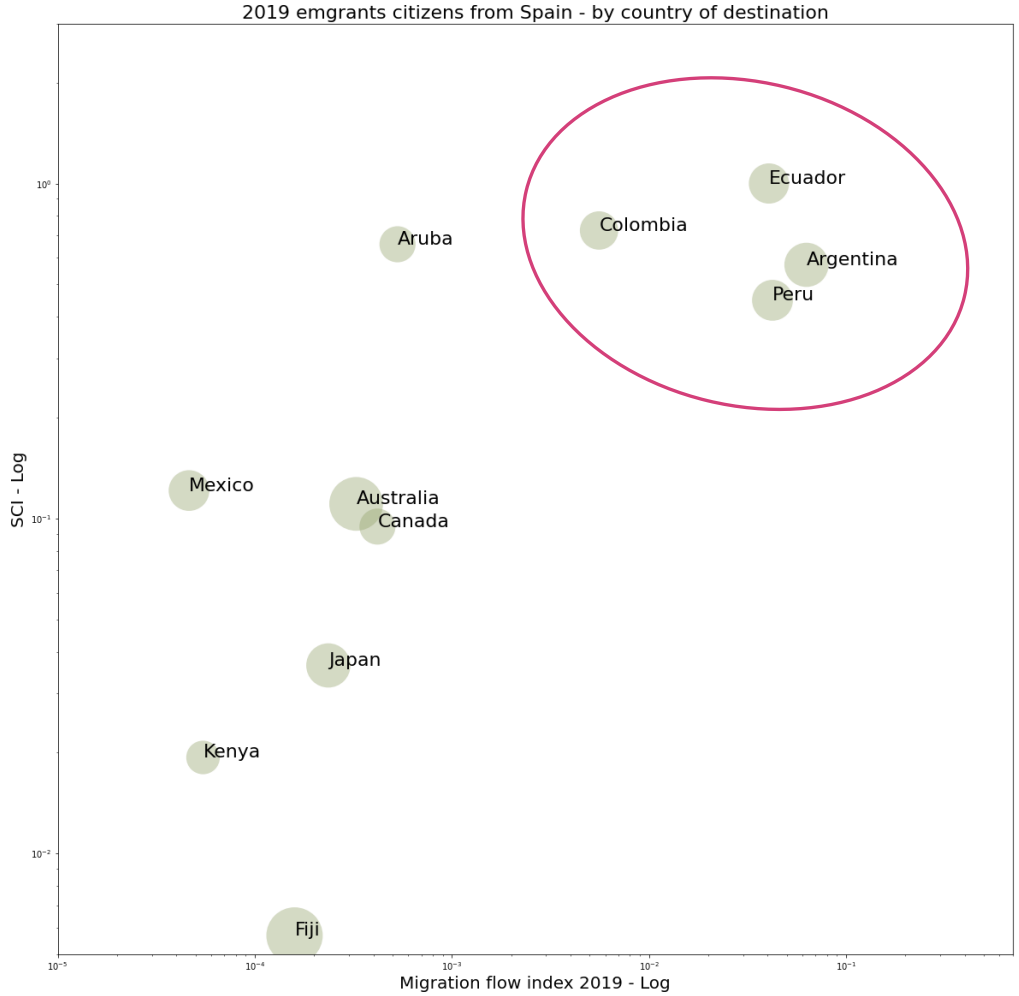
Most connected countries (Equatorial Guinea, Paraguay, Honduras, Dominican Republic, Nicaragua) do non share migration flows, but… they are Spanish colonies! Common history, language, culture, religion.

Connectivity of Spain

Top-30 countries having stronger connections with Spain

Migration flows to/from Spain

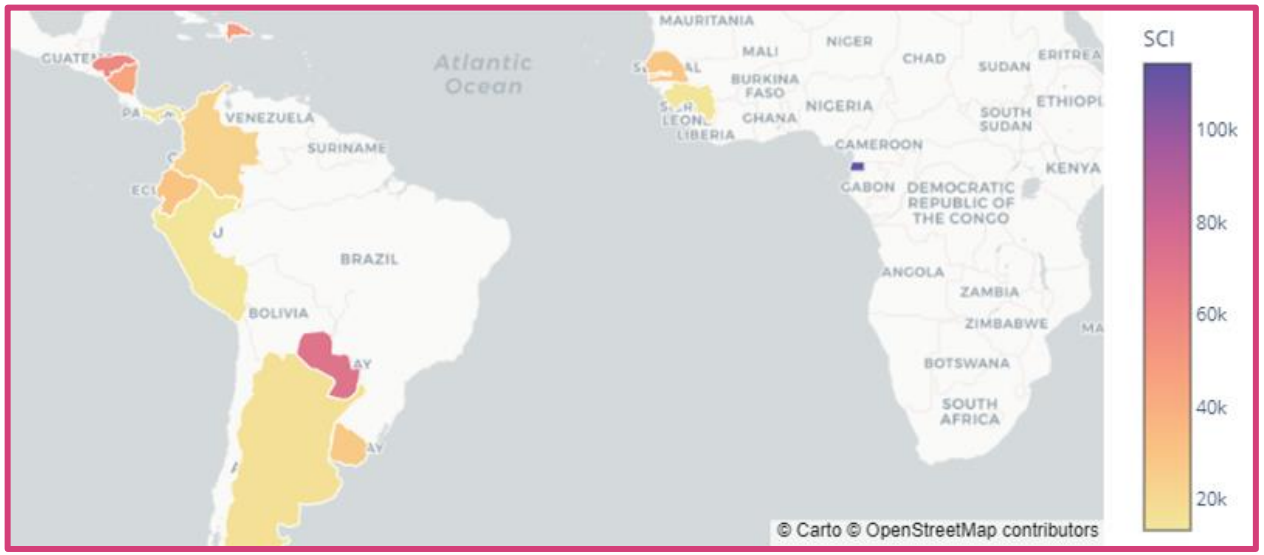
International future residence



Most connected countries (Equatorial Guinea, Paraguay, Honduras, Dominican Republic, Nicaragua) do non share migration flows, but… they are Spanish colonies! Common history, language, culture, religion.

Connectivity of Spain

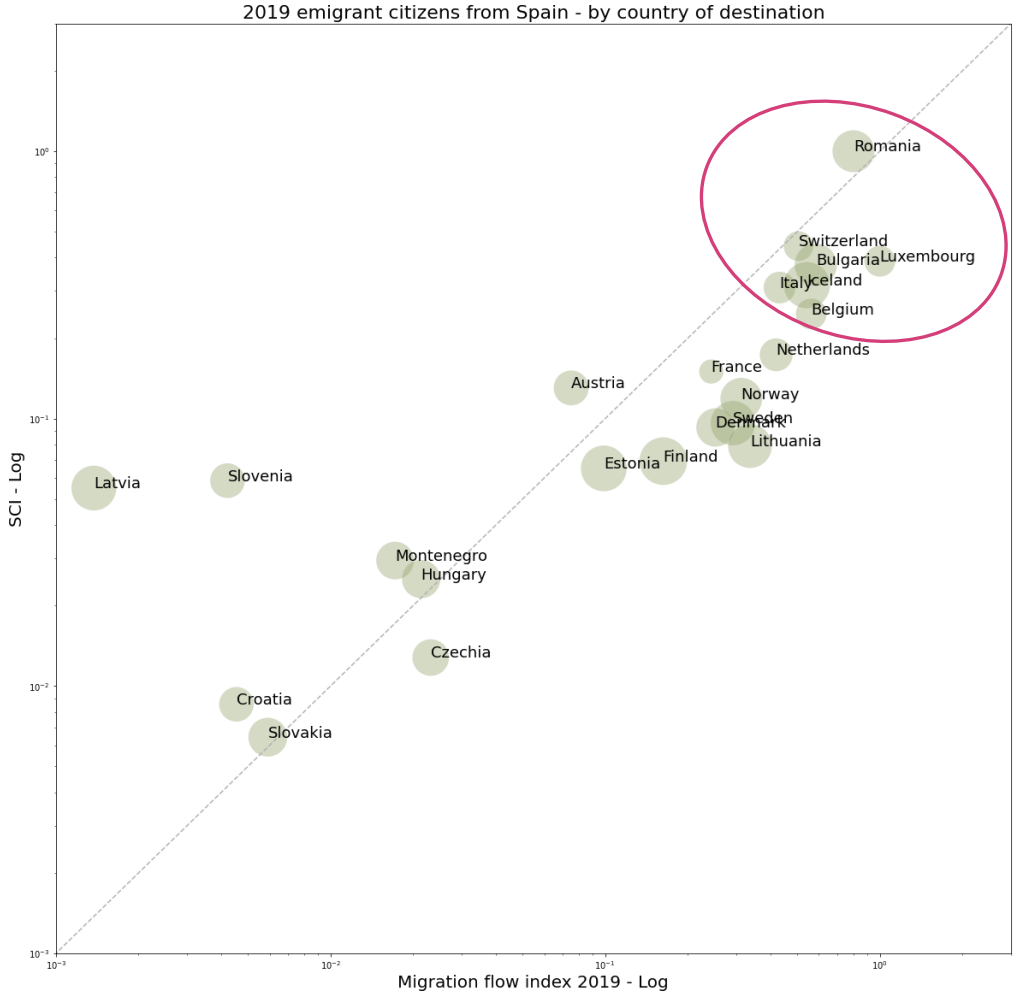
Top-30 countries having stronger connections with Spain



Colonies

Migration flows to/from Spain

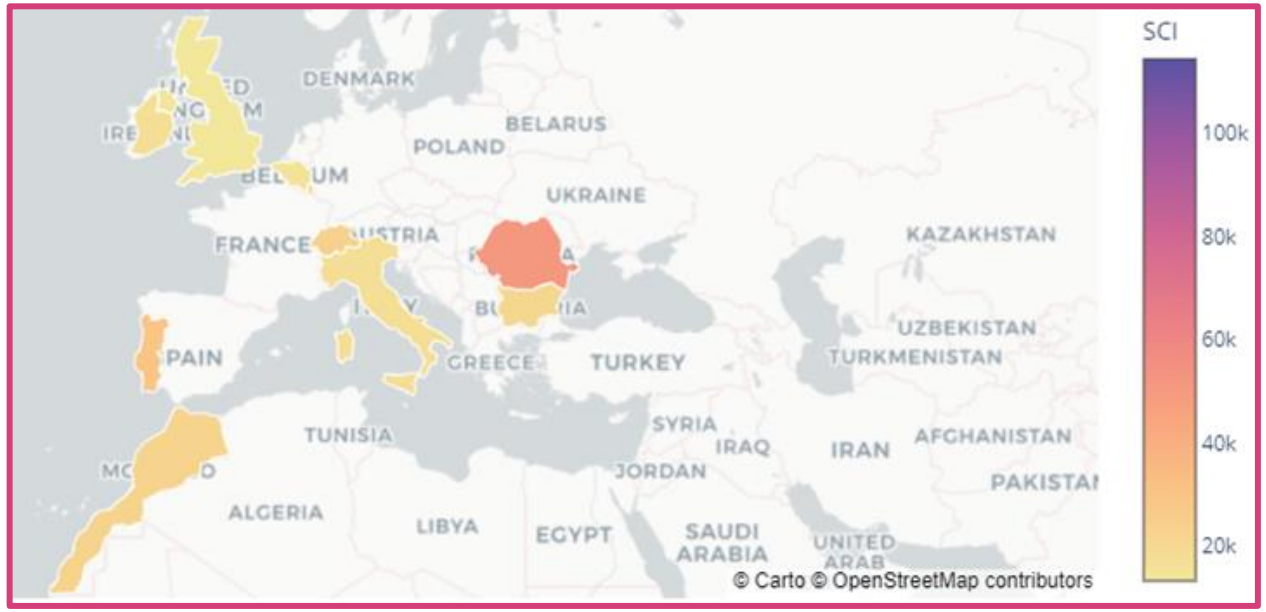
Future residence in Europe



Most connected countries (Equatorial Guinea, Paraguay, Honduras, Dominican Republic, Nicaragua) do non share migration flows, but… they are Spanish colonies! Common history, language, culture, religion.

Connectivity of Spain

Top-30 countries having stronger connections with Spain



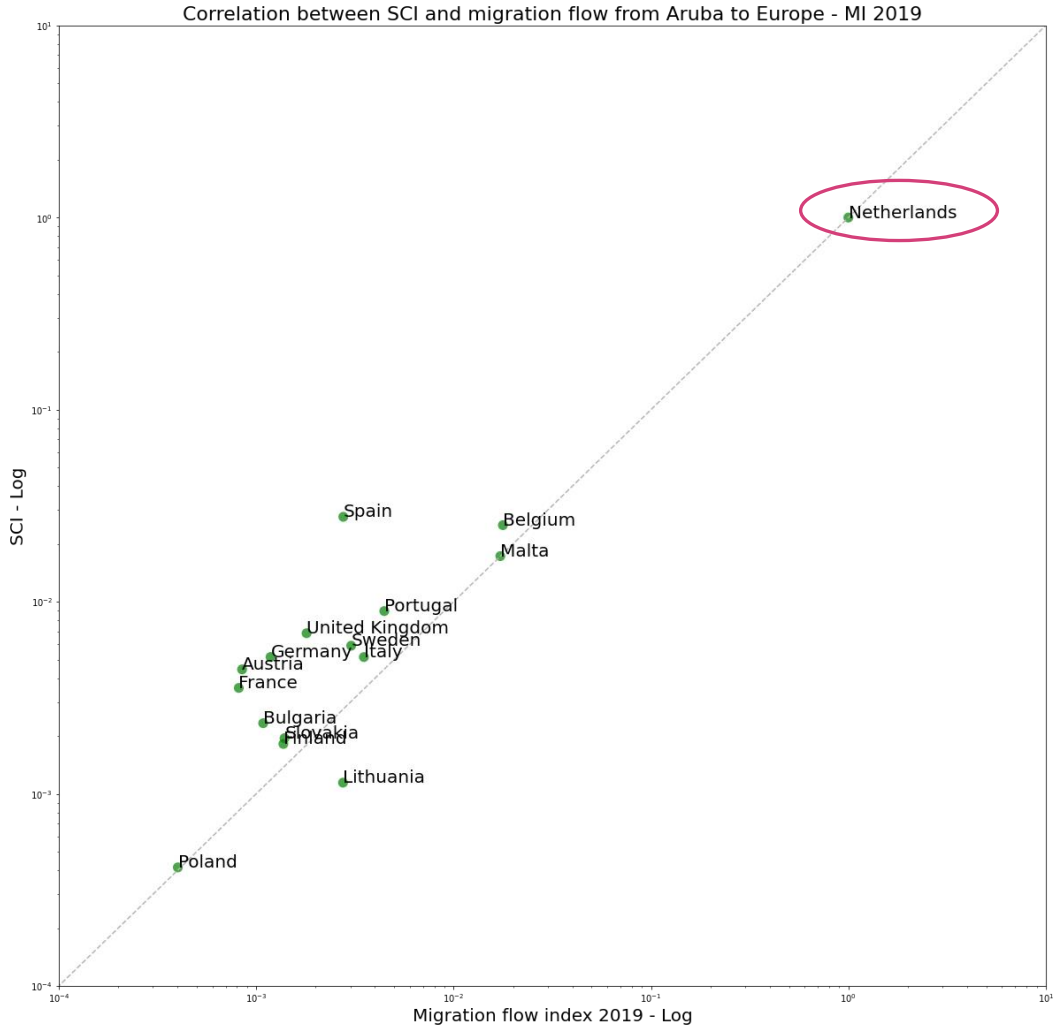
Colonies, historical background & cultural similarities

Case study: **Aruba,** Constituent Country of the Kingdom of the Netherlands (Dutch Caribbean islands)

	provenance	destination	MI_2019	sci
0	Aruba	Netherlands	70,404.40	707,901.00
1	Aruba	Belgium	1,252.75	18,170.00
2	Aruba	Malta	1,215.83	12,678.00
3	Aruba	Portugal	314.89	6,779.00
4	Aruba	Italy	248.39	4,098.00
5	Aruba	Sweden	214.21	4,629.00
6	Aruba	Spain	195.59	20,032.00
7	Aruba	Lithuania	194.86	1,262.00
8	Aruba	United Kingdom	127.75	5,303.00
9	Aruba	Slovakia	99.03	1,832.00

Top-10 European destinations for emigrating citizens from Aruba, sorted by highest values of MI 2019.

Correlation SCI-MI	
Method	Value
Pearson	1
Kendall	0.3
Spearman	0.4



Same behavior encountered for emigration from autonomous regions / territories and their respective sovereign state. Unfortunately we do not have connectedness data for them.

From Åland Islands (autonomous region of Finland) and from Faroe Islands (autonomous territory within the Kingdom of Denmark)

T5.1 Summing up

- Relation between Facebook connectivity and migration trends has been proved.
- Look at it as a **starting point**.
- Challenges:
  - matching, uniform and standardize different nomenclatures
  - characterize outliers
- New results in literature: being finalized and concretized in a paper.

Call for contribution!

Next steps

HumMingBird

Next step: **nowcasting.**

- ML model (regression) for **migration prediction** based on Facebook connectivity

Retrieve and include new relevant features

- Country area (km2), average per capita income (including GAP for pairs or countries), average education leve employment rate, religion, language, common historical background, other cultural indicators.

Timeseries analysis

- Include older data (prior to 2018)

More qualitative analyses

- Include migration stocks
- By age groups and sex (particular focus: women and children)
- But also: returners, highly skilled migration, brain gain trend. Refugees, migration corridors.

Dataset upload and publication

- Put European and global data together
- Integrate with other sources (Migration Data Portal, OECD.Stat)
- Datapaper publication