

Big Data and Migration: Limits and Perspectives

Stefano M. Iacus

University of Milan
VOICES from the Blogs
R Foundation for Statistical Computing
ISTAT: Big Data Committee

(on-going work with Y. Teocharis)

Big Data and Alternative Data Sources on Migration:
from case studies to policy support, JRC-ISPRA, 30 November 2017

Why Social Media and Refugee?





SHARE



SHARE
1439



TWEET



PIN



COMMENT
45



EMAIL

ALESSANDRA RAM GEAR 12.05.15 7:00 AM

SMARTPHONES BRING SOLACE AND AID TO DESPERATE REFUGEES



Why Social Media and Refugee?

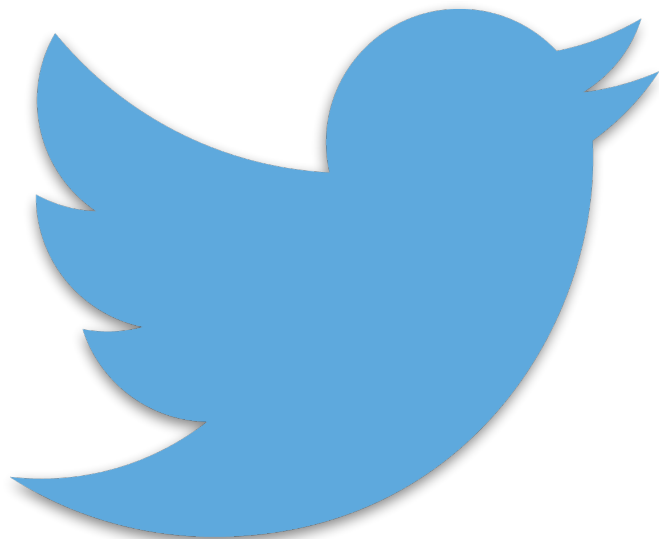
- **Refugees*** are **coming from highly wired/connected countries**, with youths that are experts in ICT use and high levels of social media use (Howard, 2011; Howard & Hussain, 2014)
- **Smartphones** are the **most important items in** most refugees' **luggage**. About the only things available to them to **keep in touch with people at home**
- The best possible **tool** in their hands to **document the conditions they live in** and their struggle for physical survival
- Help people create and share their own narratives through systematic visual and textual documentation
- Allow them to present themselves as **human beings rather than as “others” or “hostiles”** or whatever else

*** At least those from the Syrian crisis**

Part of Twitter & Instagram have geo-reference meta data

For **Twitter**, this proportion is around 1% to 3% of the total accouts

```
{
  "geo": {
    "type": "Point",
    "coordinates": [40.0160921, -105.2812196]
  },
  "coordinates": {
    "type": "Point",
    "coordinates": [-105.2812196, 40.0160921]
  }
}
```



Do we have enough data then?

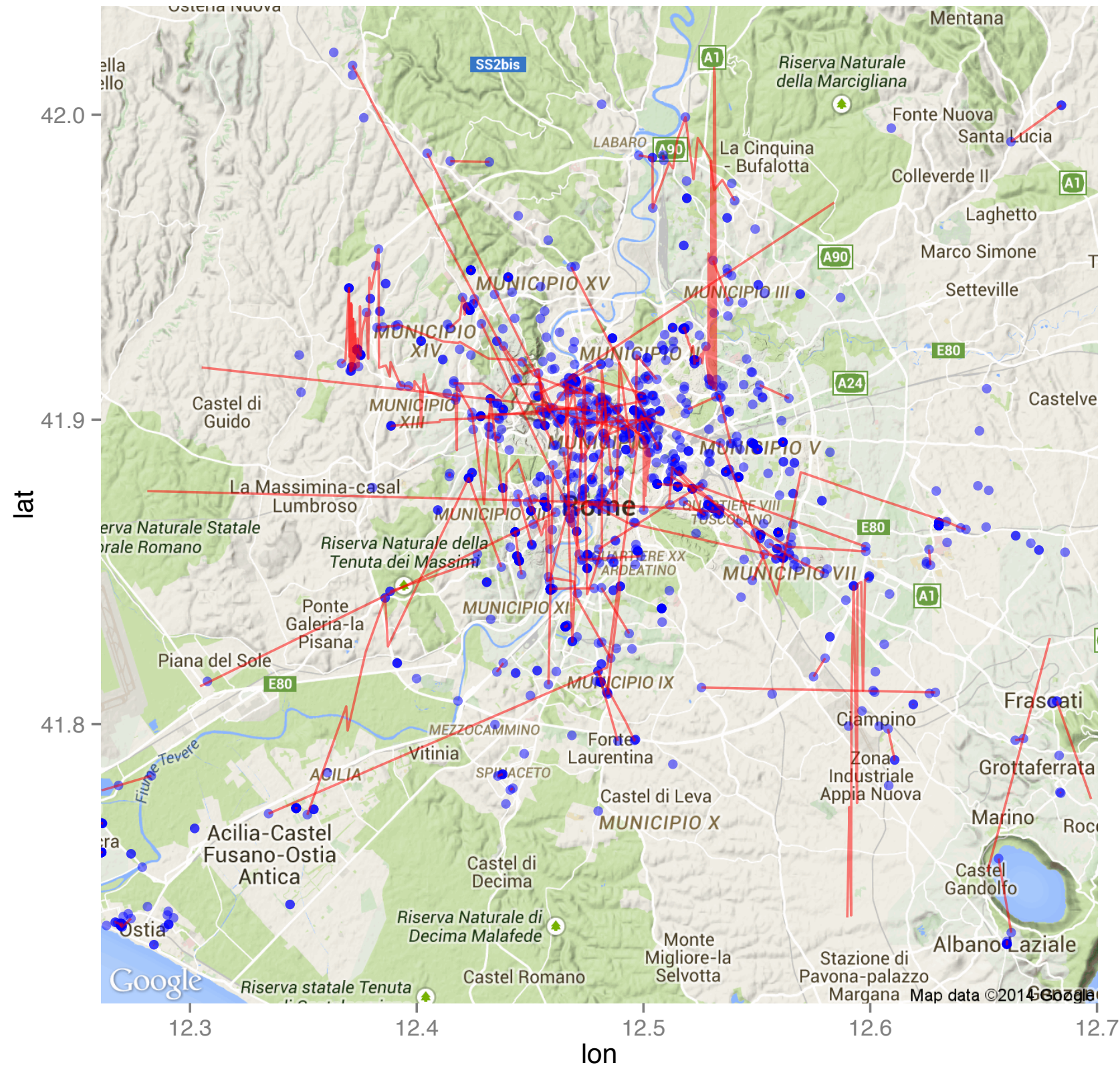
Part of Twitter & Instagram have geo-reference meta data

For **Instagram**, this proportion grows to 30%

```
{  
  "data": [{  
    "id": "788029",  
    "latitude": 48.858844300000001,  
    "longitude": 2.2943506,  
    "name": "Eiffel Tower, Paris"  
  }]  
}
```



Application from tourism study

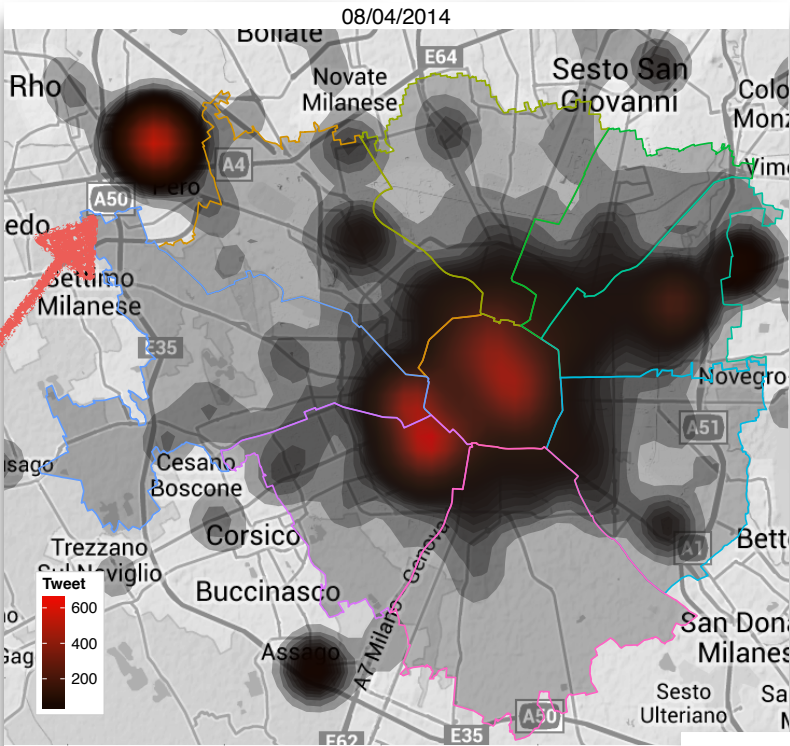


Movements
in a day in
Rome

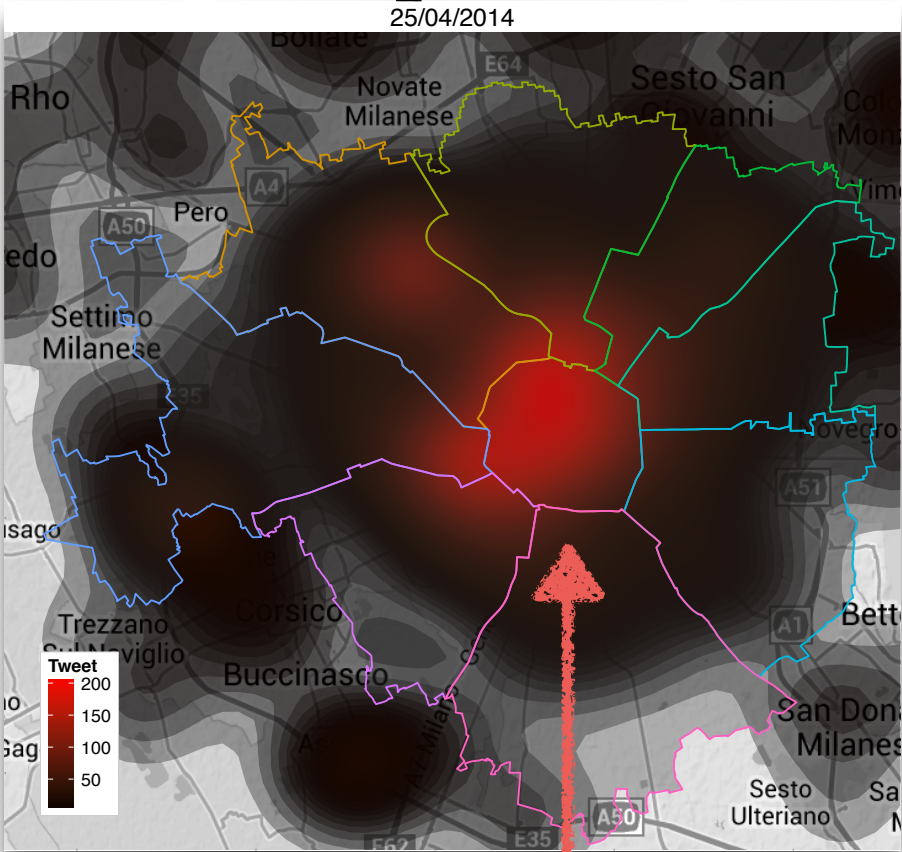
Density along the year in Milan

Salone Mobile

Fiera Milano Rho

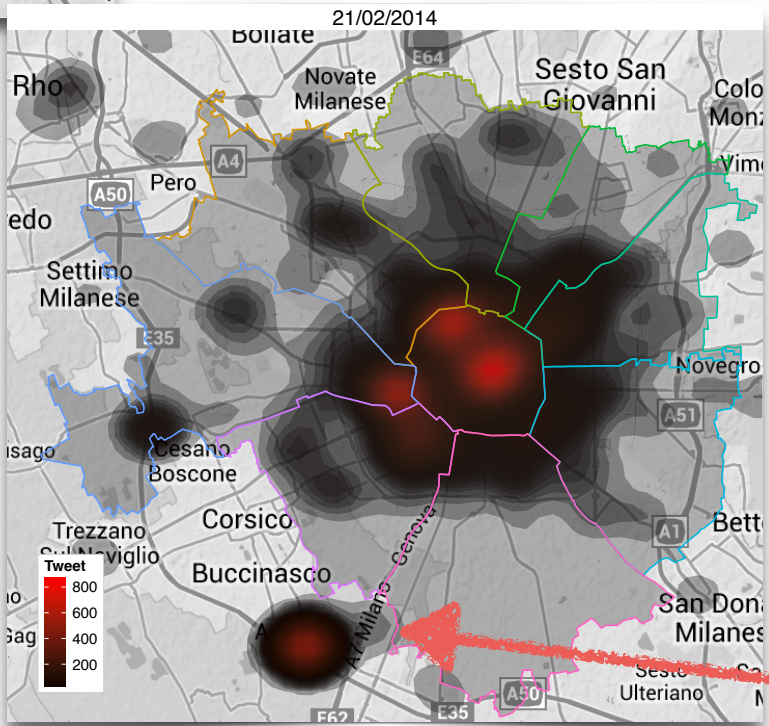


April 25th



Duomo Square

Concert

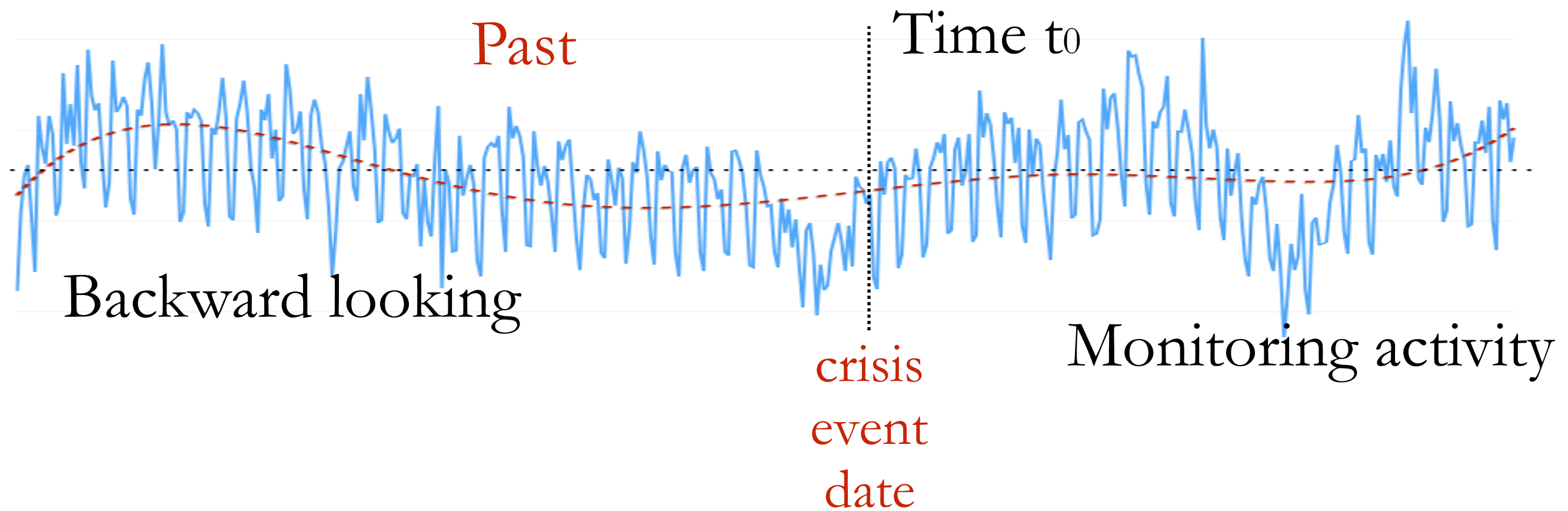


Forum Assago

Travels of italian Twitter accounts through 2012-2016



Travel in time



The ongoing Refugee project



About 5000 Instagram posts limited to the Idomeni camp area in Greece

The ongoing Syrian Refugee project

About 5000 Instagram posts limited to the Idomeni camp area
in Greece

Period 11-21 Feb 2016

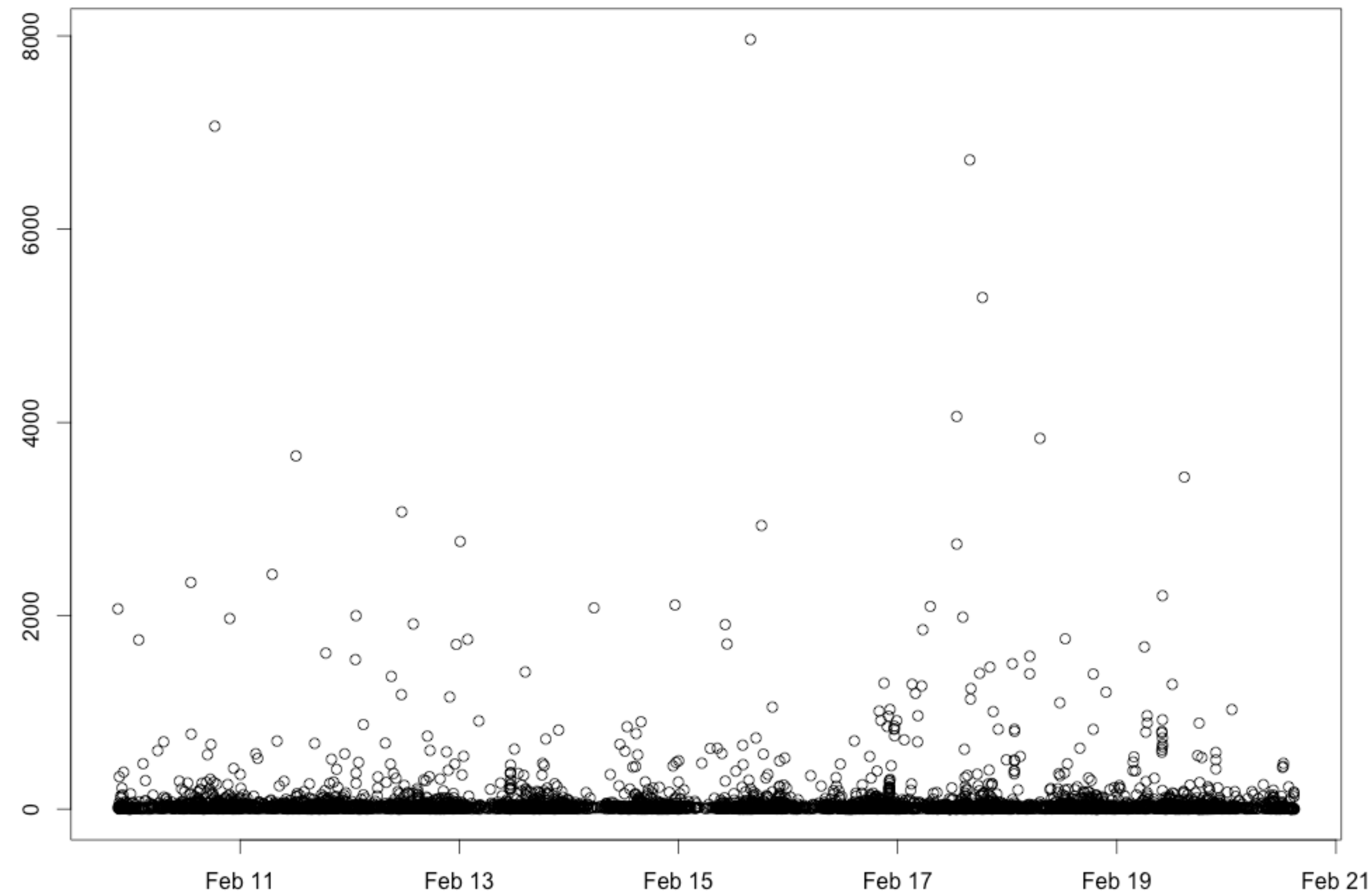
Plan: track the accounts generating these posts a year later

Issue: Instagram has severely restricted the API usage in late April 2016 and this makes this repository **very valuable but**
data analysis **extremely hard!**

The ongoing Syrian Refugee project



Activity around the 5000 posts



The ongoing Syrian Refugee project

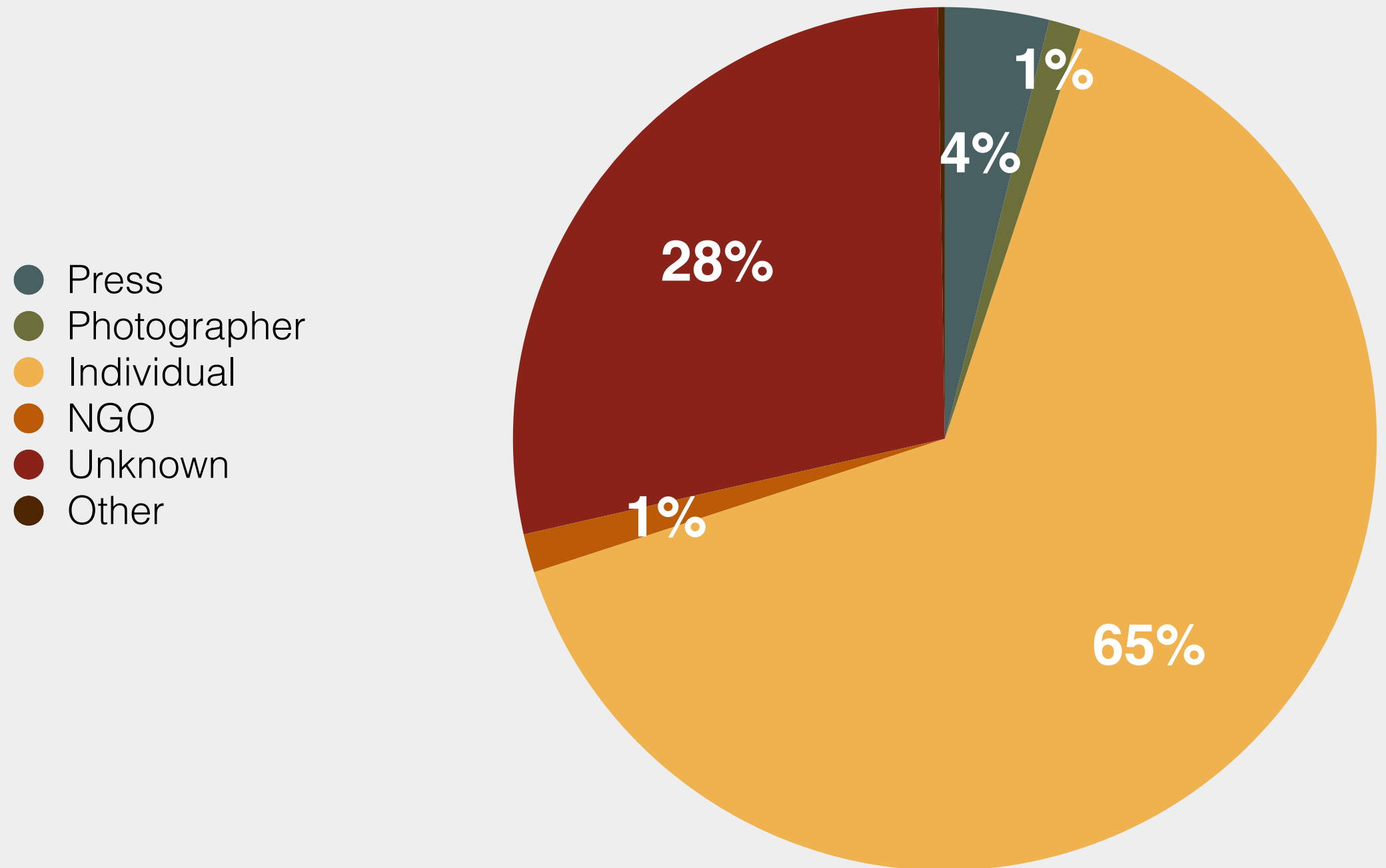
1. Extract the Instagram accounts present in the data collection from **Idomeni** camp (Greece) ✓
2. **Follow** them in **future** (i.e. present): are they still in Greece? Or they moved around Europe? ✓
3. Check if **before** the **crisis** the accounts were posting from, e.g., Syria, to test for “real” (suspect) refugee ✗
4. confirm the analysis by manually looking at the profiles ✓

The ongoing Syrian Refugee project

1. We proceed by **manual inspection** of **3K** Instagram profiles looking at last three months (Sep-Nov 2017) of their timeline to determine: country of destination, last city visited, type of account (Individual/NGO/Press/etc) and Refugee status by contacting the users as followers of their accounts and/or asking for permission to see their images
2. We failed to get permission in 28% of the cases and those are likely to be some of the interesting accounts
3. Proportion of true Refugee left in the sample very low: 1% of clear refugee status and 30% of “difficult to assess” cases

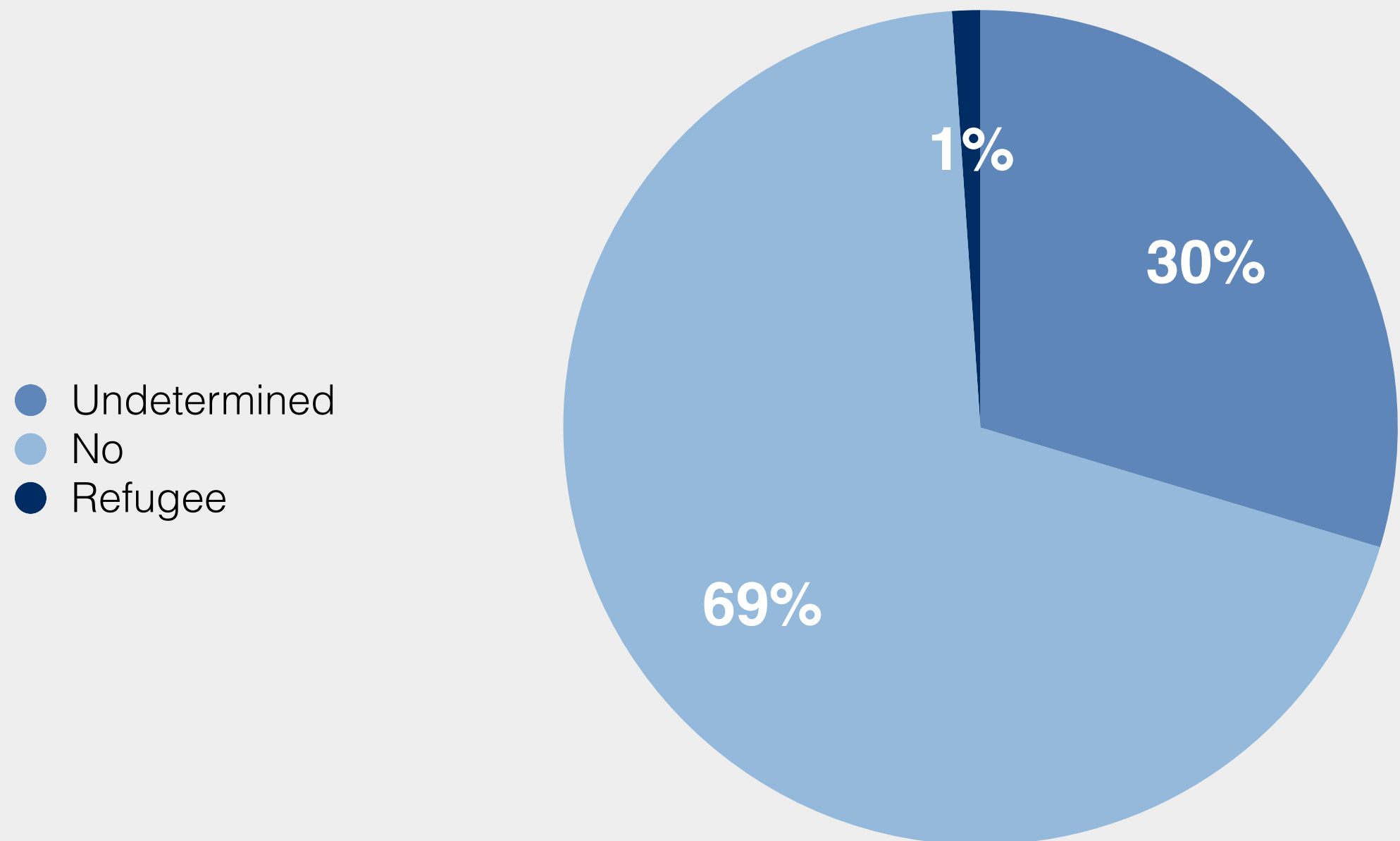
Current results of the analysis

Number of accounts: 3206, mostly personal accounts



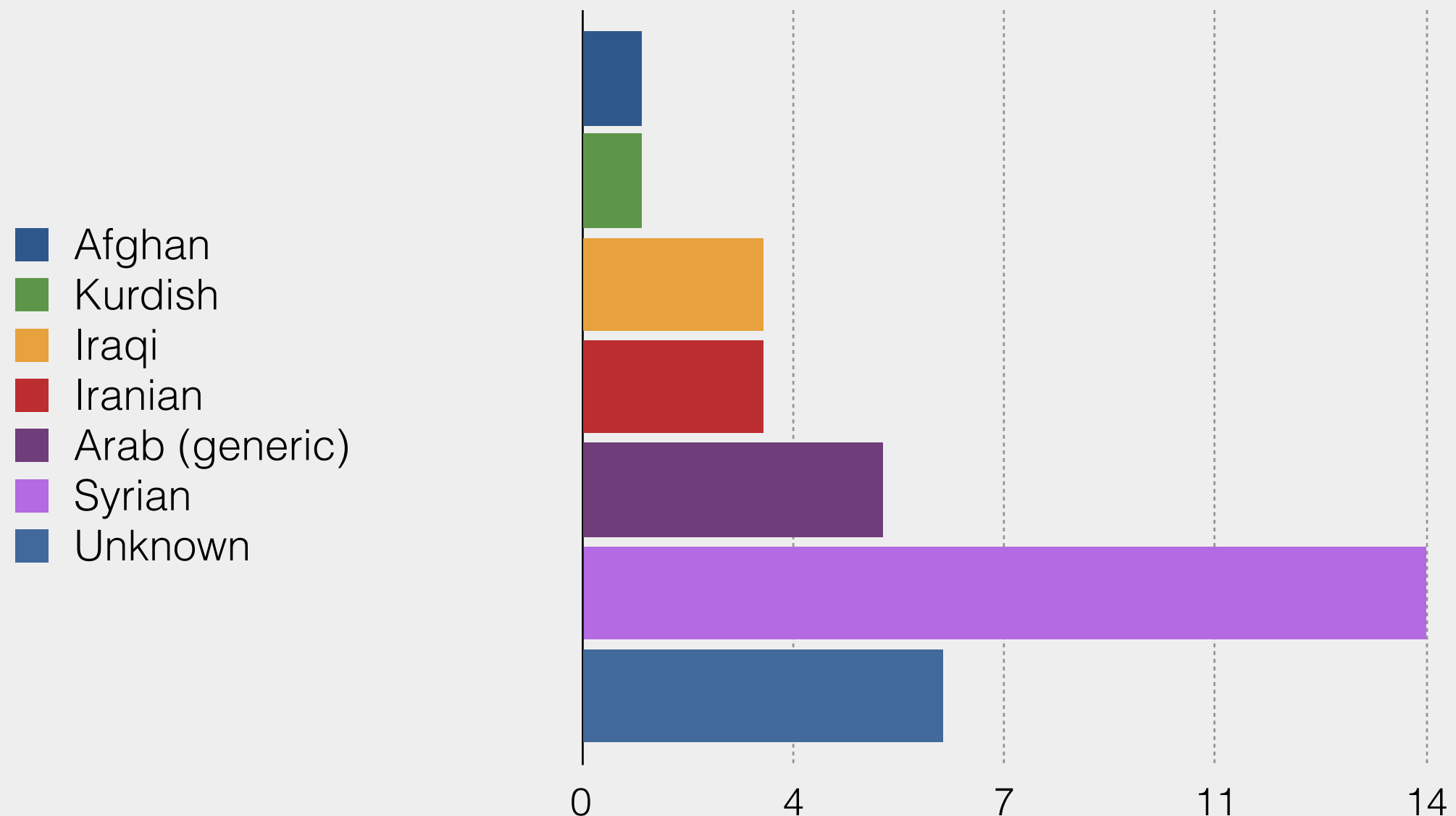
Current results of the analysis

Number of accounts: 3206,
very few assessed refugees (1%) and 30% of uncertain cases



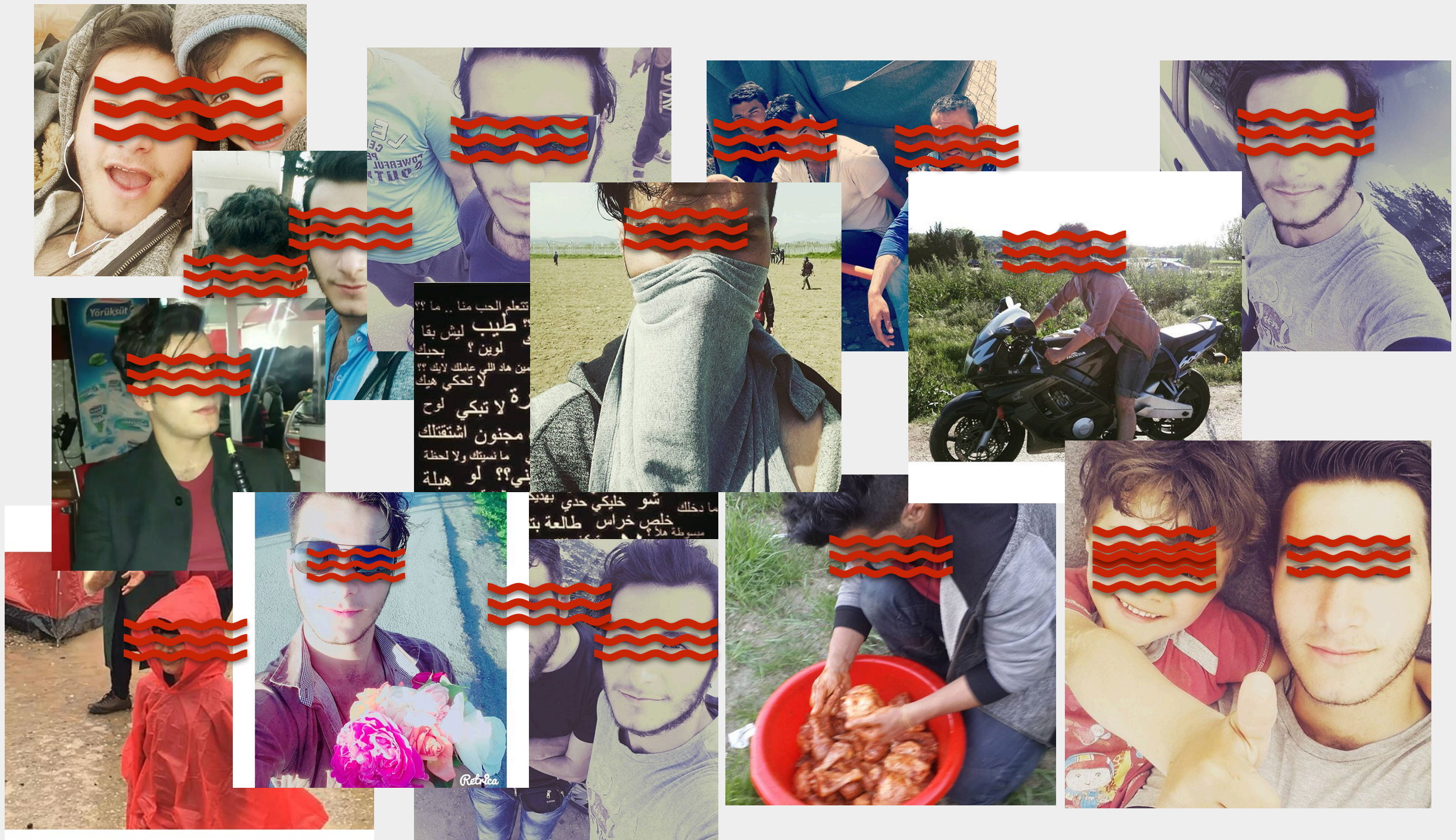
Current results of the analysis

Syrian are the majority though some other nationalities exist



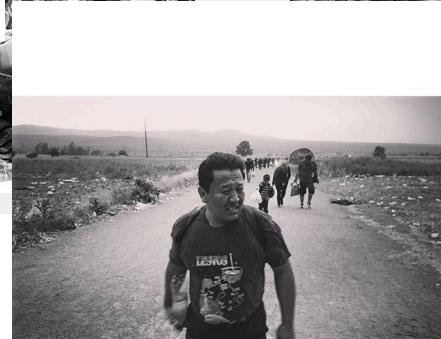
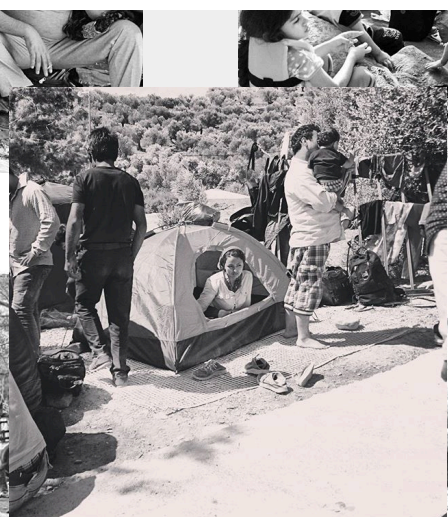
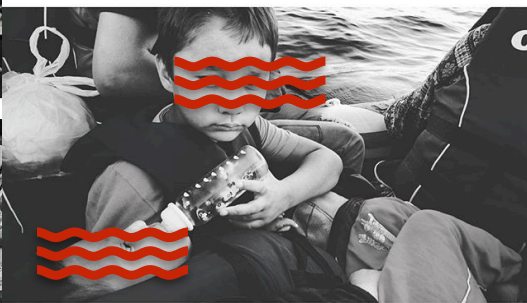
Current results of the analysis

What did refugees share: **personal life**



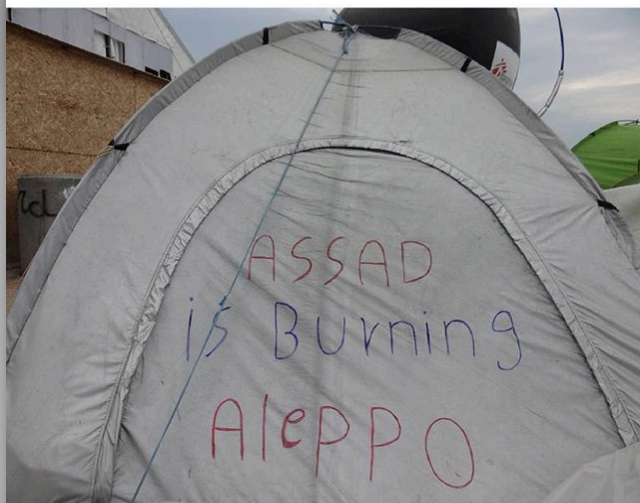
Current results of the analysis

What did refugees share: **camp life**



Current results of the analysis

What did refugees share: **politics/protest**



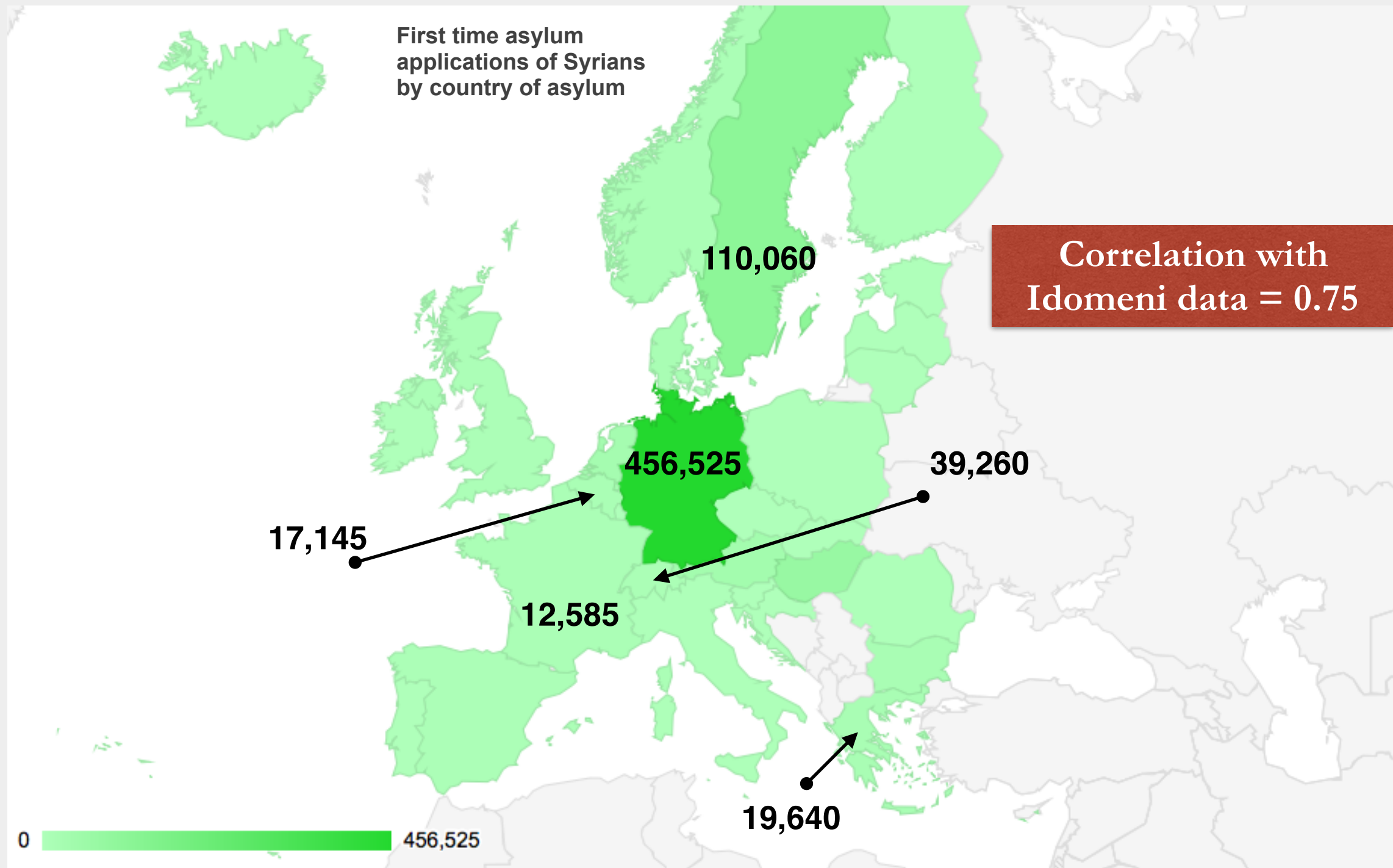
Current results of the analysis

Flows for fully identified refugees

Origin vs Destination	Germany	Greece	Austria	Sweden	Belgium	Iran	France	Turkey	Total
Syrian	8	3	1	1				1	14
Arab (generic)	4			1					5
Iranian	2			1					3
Iraqi	1		1		1				3
Kurdish	1		1				1		3
Afghan						1			1
Unknown	2	4							6
Total	18	7	4	2	1	1	1	1	35

Current results of the analysis

Flows for fully identified refugees: spurious correlation?



Cumulative flows over the period 2010 – 2016.

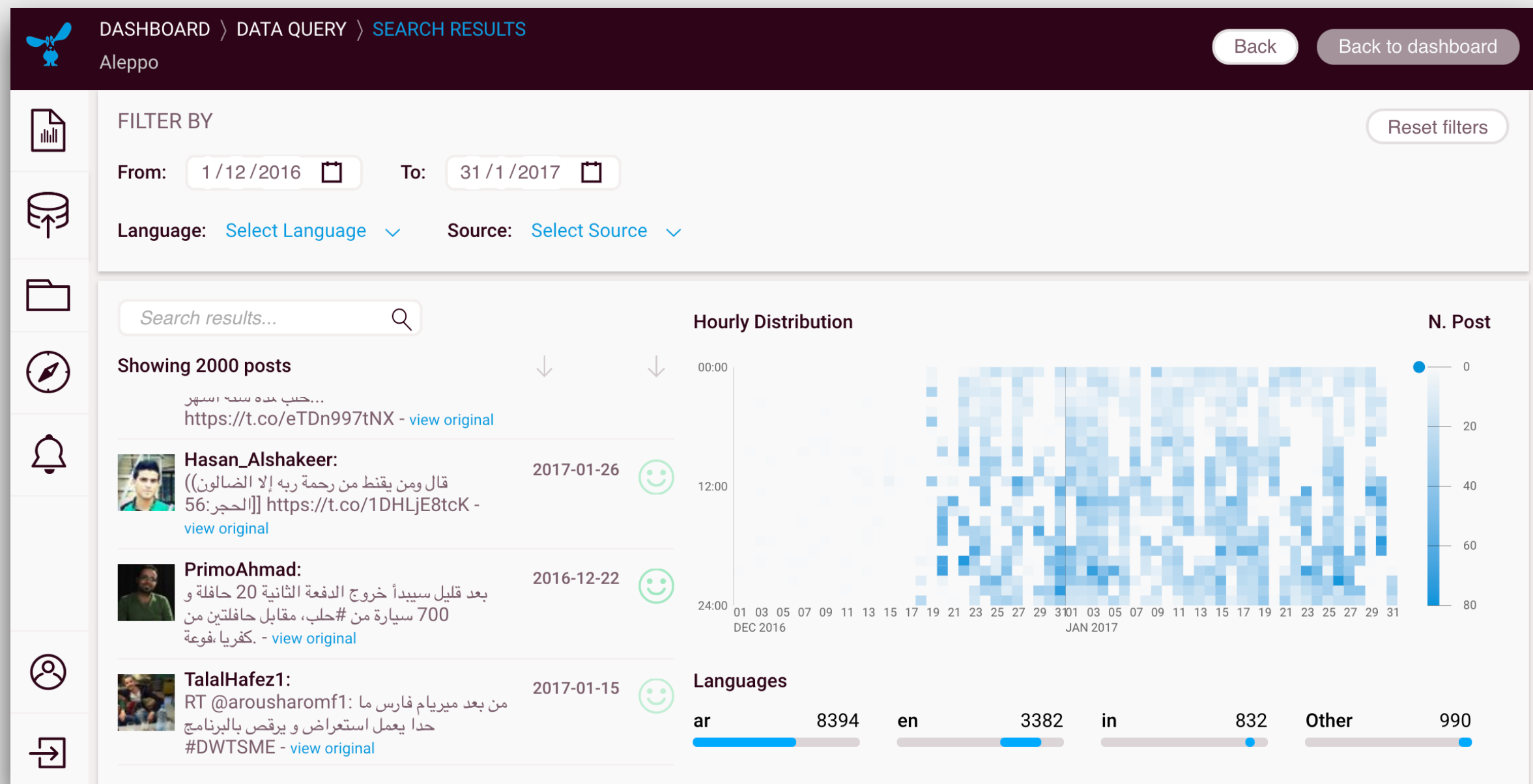
Source : Elaboration by <http://syrianrefugees.eu/> based on Eurostat (EU28, Norway and Switzerland).

Evidence so far

- social media data are **biased but** somehow **capture the trends**
- **noise is predominant** in social media data, **but signal exists**
- social media platform may abruptly change or close the API's and therefore invalidate workflow of analysis/monitoring
- **privacy** of social media users is a issue (in terms of official statistics)
- still lots of **ad hoc** analysis

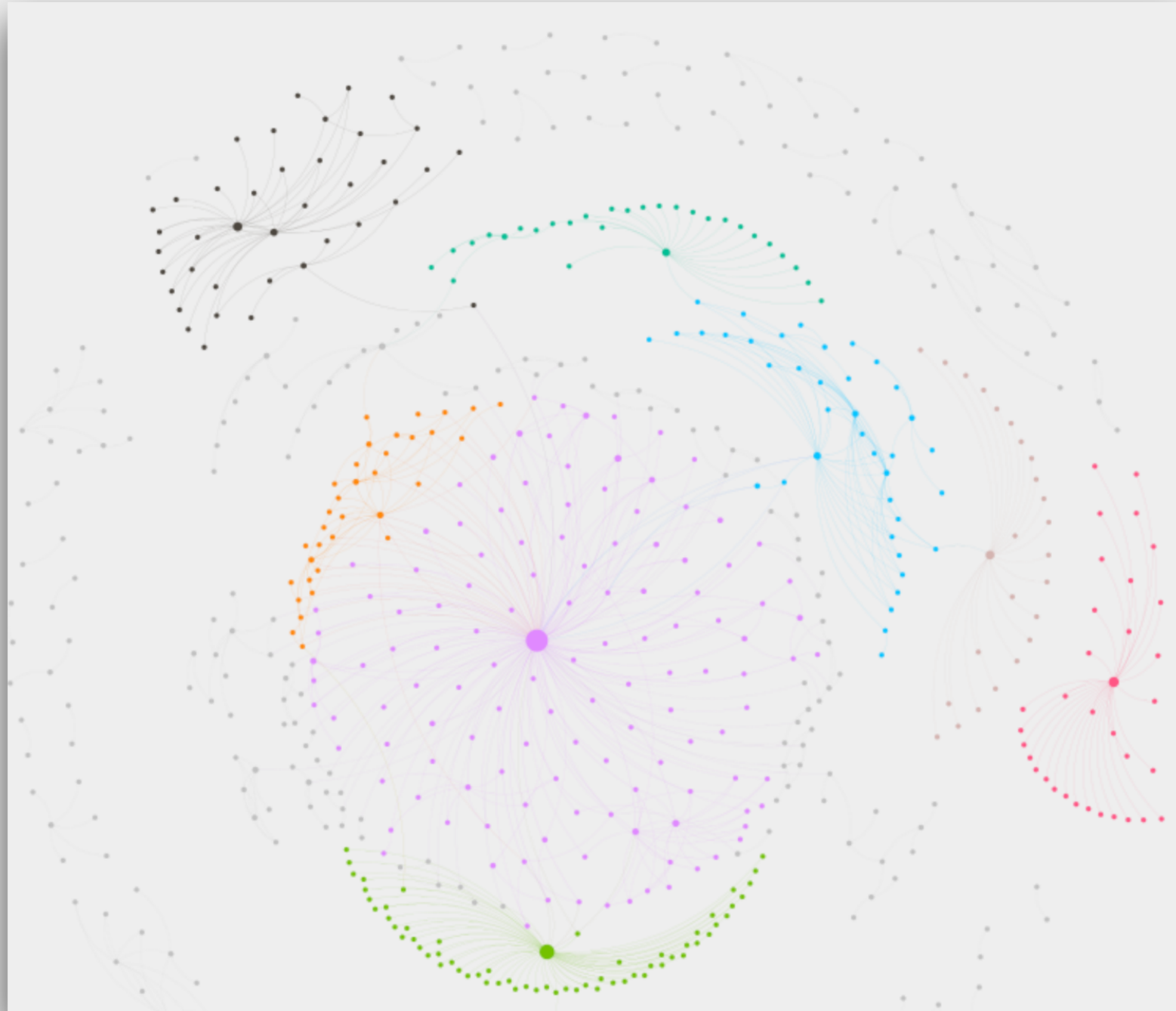
Parallel Project

We are monitoring all Twitter posts from Aleppo since December 2016. This time about 1 Million posts. Twitter API still open!



We will produce a migration map by looking at the accounts after one year and cross the results with other Internet data and official statistics.

Can we also exploit the information coming from networks?
Which are the hubs of information?
Would monitoring the hubs help in forecasting the flows?



Contacts

Twitter @iacus

WWW <http://www.voices-int.com>

Mail stefano.iacus@unimi.it

