

# Understanding International Migration



## Using Tensor Factorization

Hieu Nguyen, Kiran Garimella

Aalto University

{hieu.nguyen,kiran.garimella}@aalto.fi

Dataset and demo available on at <https://users.ics.aalto.fi/kiran/migrationTwitter>

### 1. Motivation

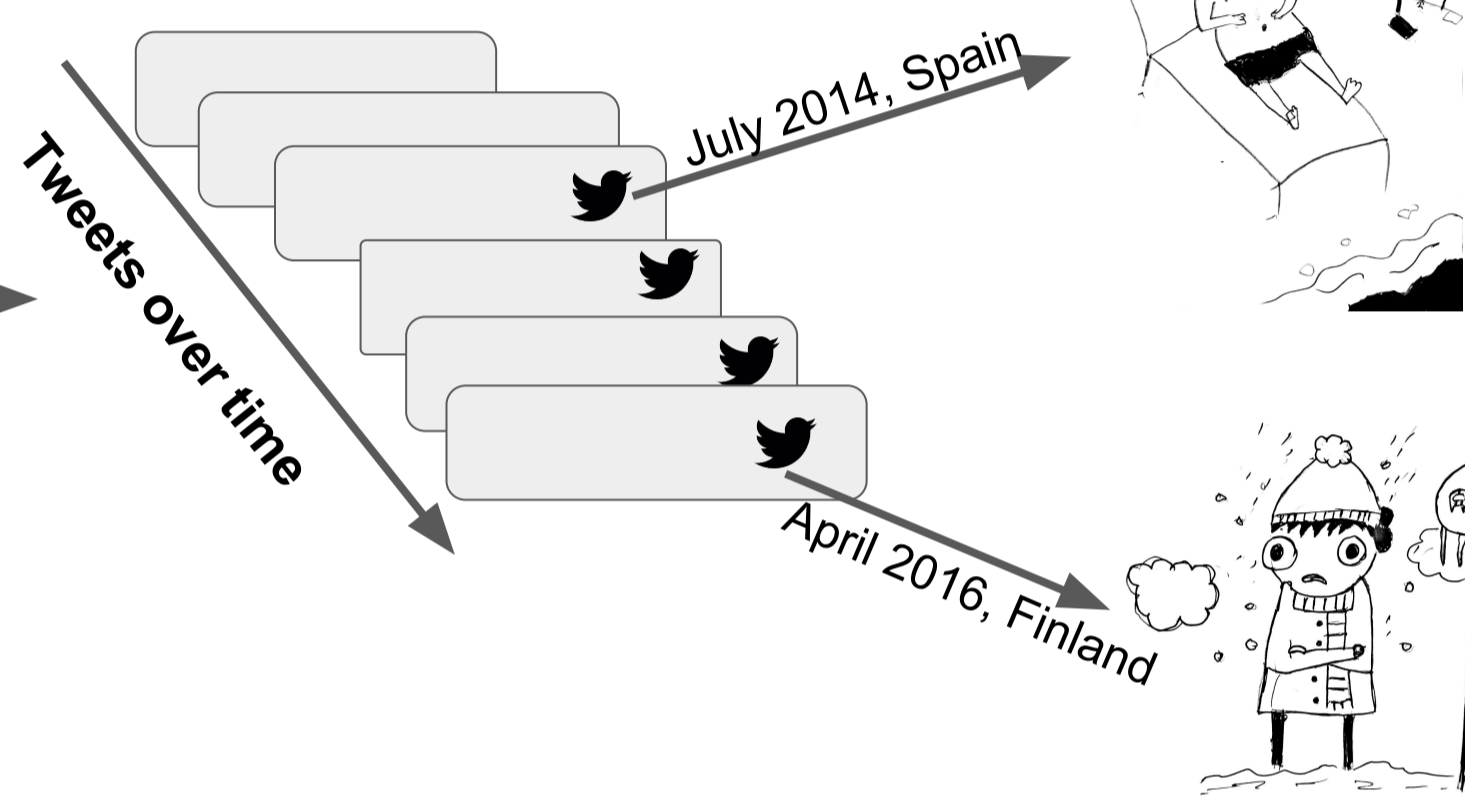
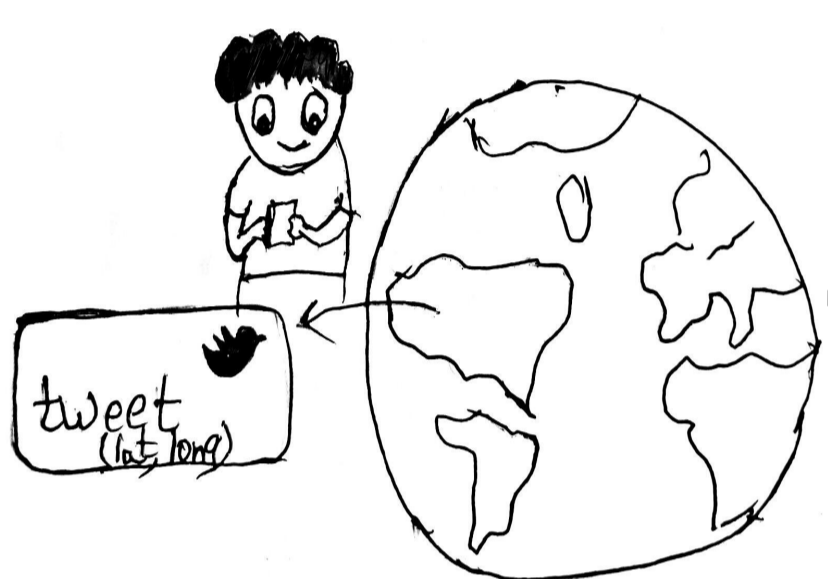
- Understanding human migration is important for social scientists, demographers and governments.
- Social media gives unobtrusive access to migration data.

### 2. Dataset

- Geo tagged tweets
- Over 5 years
- 138M tweets
- 450k users

### 3. Pipeline

#### Geo-coded tweets



#### Aggregate across users

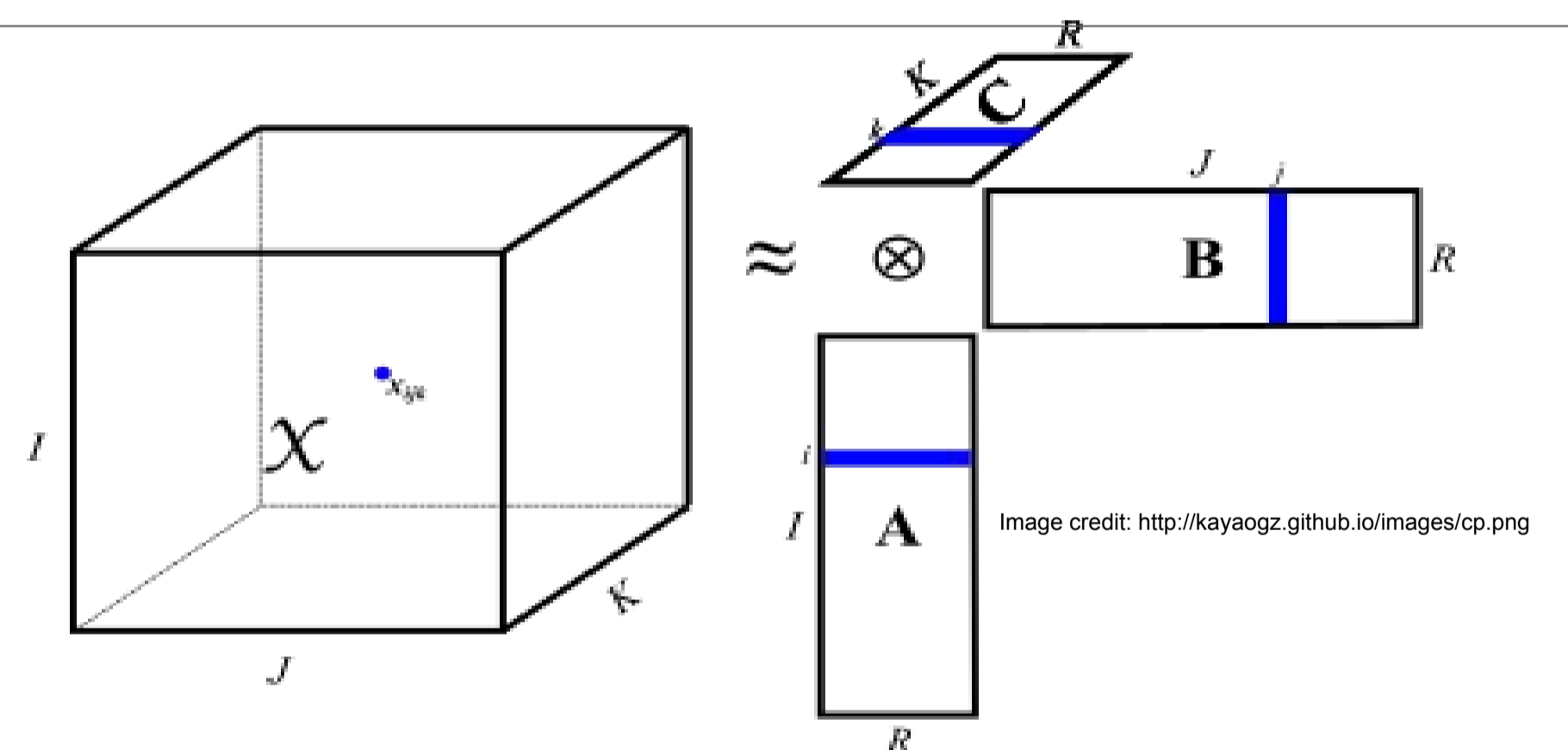
```
@user1, Jan 2014, ES
@user1, Jul 2014, ES
...
@user1, Aug 2015, ES ★
@user1, Jan 2016, FI
@user1, Apr 2016, FI
...
@user2, Jan 2012, IN
...
@user2, Jul 2014, IN
@user2, Aug 2014, IN ★
@user2, Sep 2014, US
@user2, Dec 2014, US
@user2, Jan 2015, US
```

#### Tensor construction

```
source,dest,month,count
ES, FI, Jan 2014, 2032
IN, US, Apr 2015, 1011
...
US, IN, Apr 2015, 322
MX, US, Apr 2015, 1249
....
```

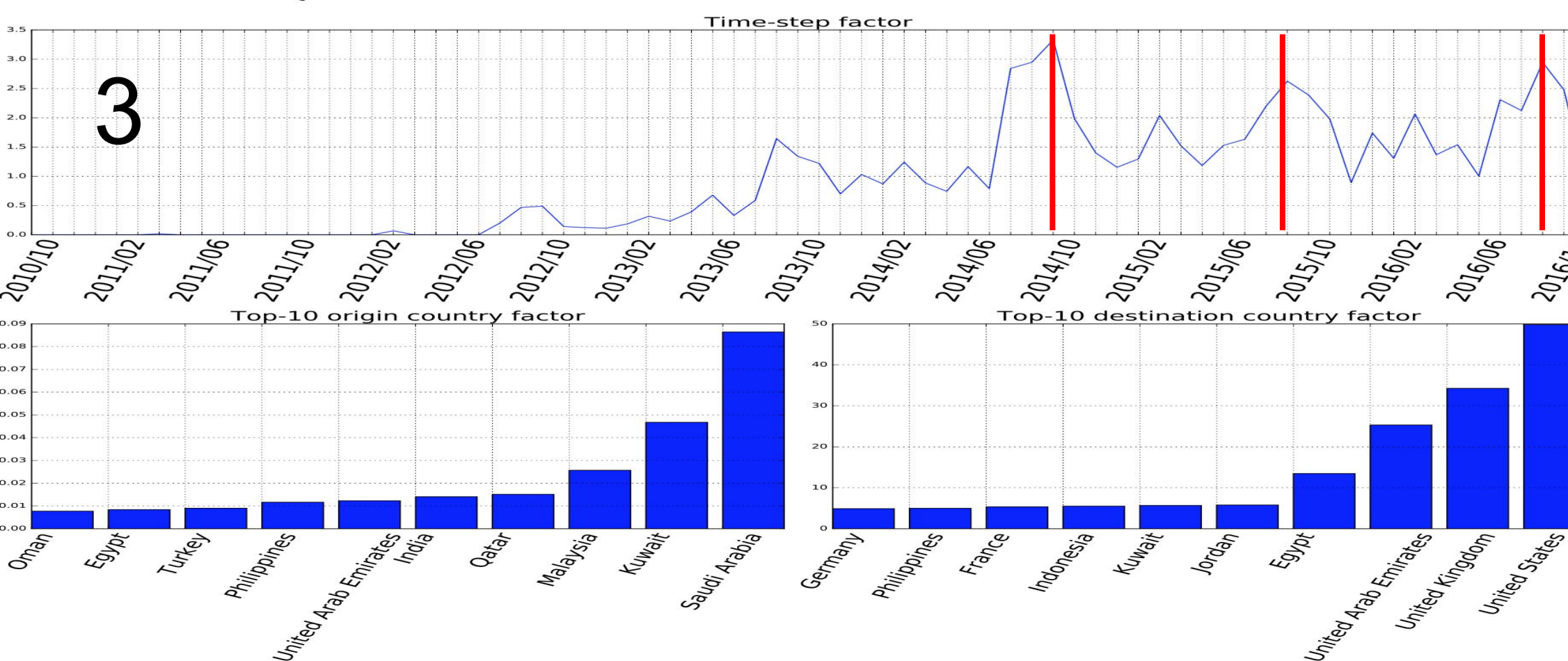
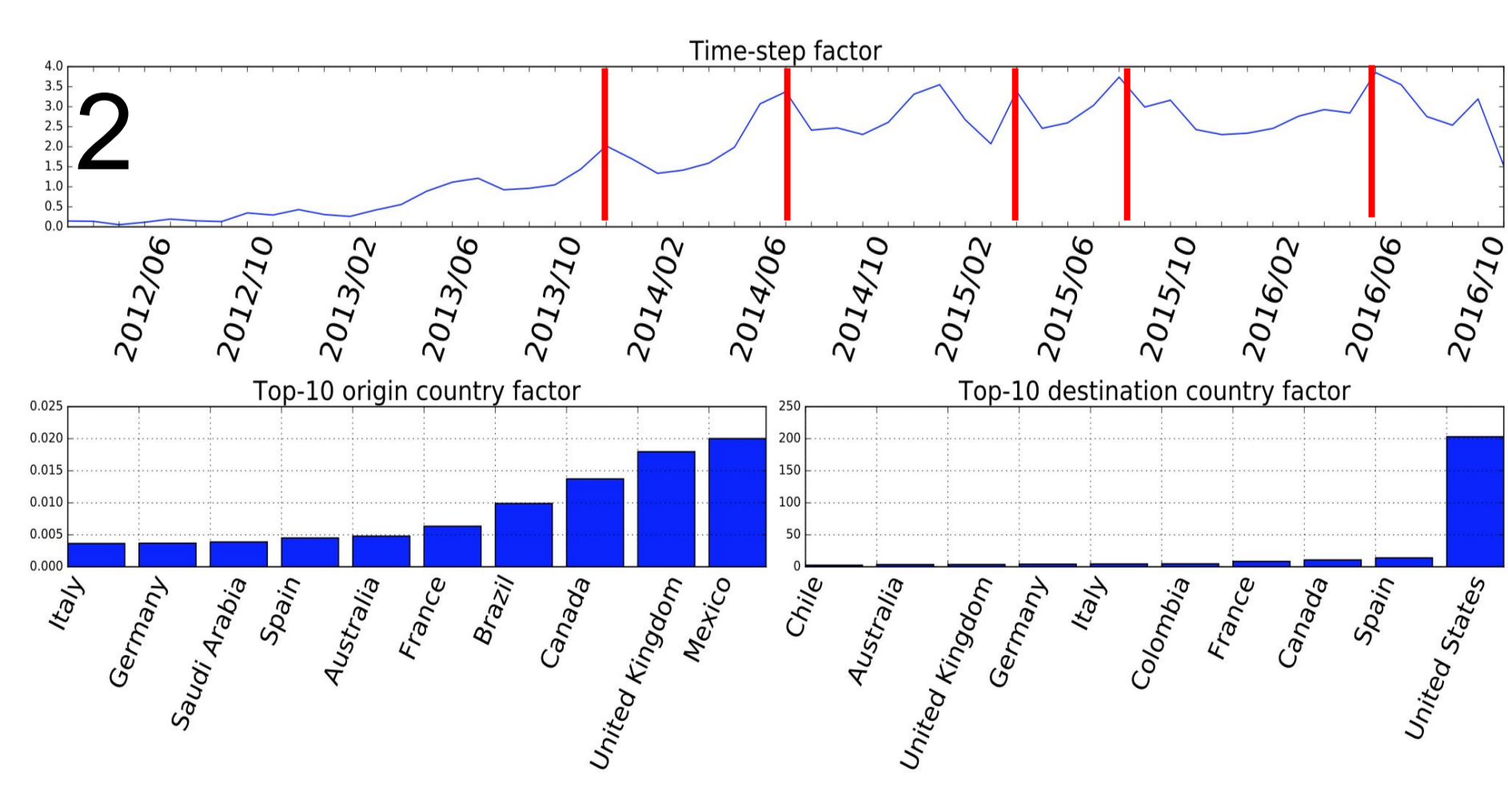
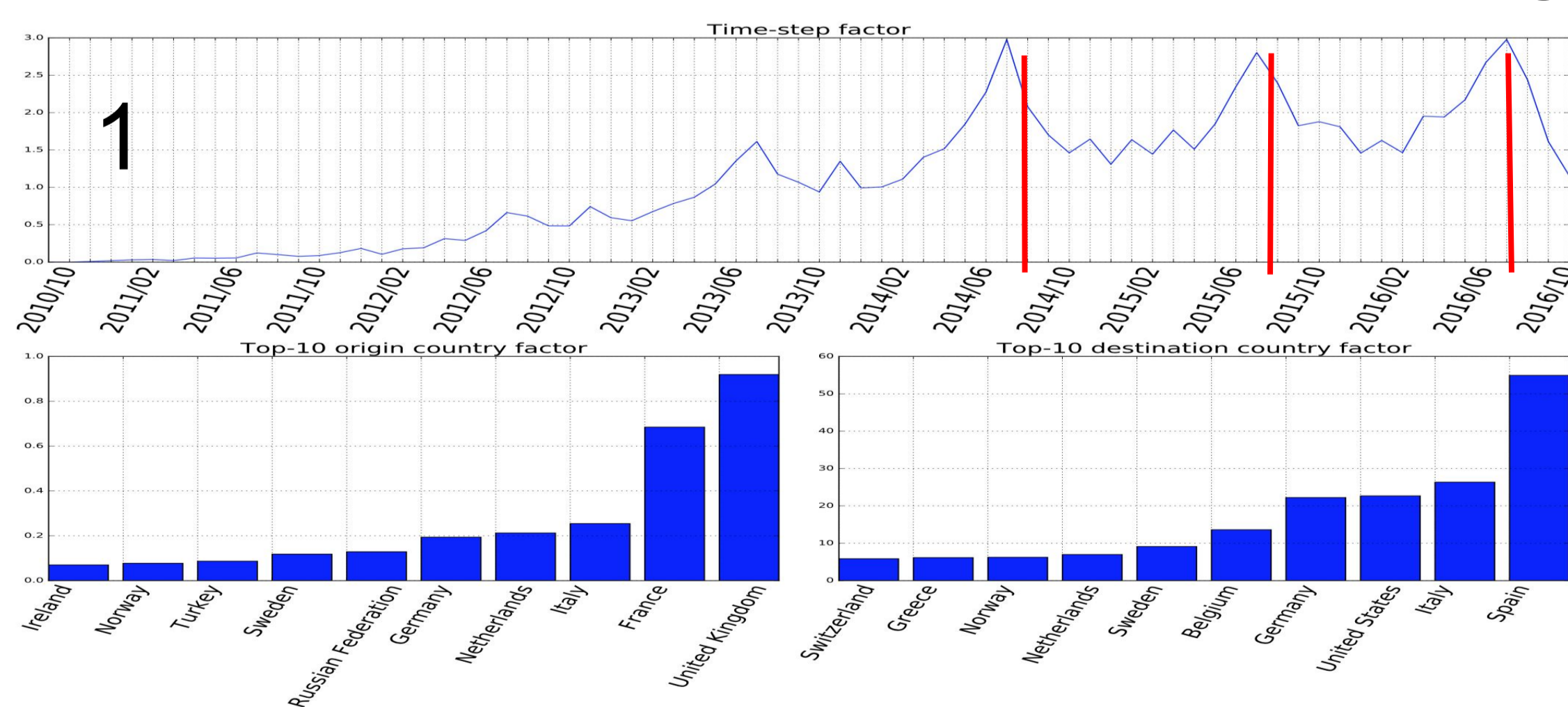
### 4. Tensor factorization

- CP decomposition
- Input: 3 mode tensor
  - Origin x destination x time step (month)
  - Size: 228 x 228 x 74
  - Parameter  $k$
- Output: 3 low rank factor matrices
  - Origin matrix
  - Destination matrix
  - Timestep matrix



- Look for outliers in the time factor

### 5. Findings



### 6. Conclusions

- Tensor decomposition shows potential
- More modes (e.g. tweet text) can be incorporated
- Biases in data not considered

#### Acknowledgements

This work has been supported by the Academy of Finland project "Nestor" (286211) and the EC H2020 RIA project "SoBigData" (654024).