Crowd forecast using mobile phone data analysis

Twitter communities in Belgium: does space matter ?

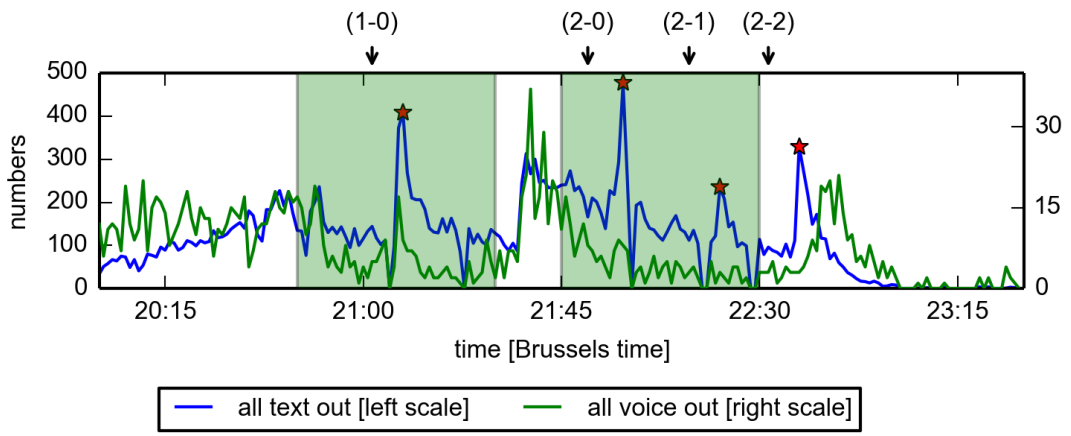
Christophe Cloquet

Université Catholique de Louvain (Belgium)

Short term crowd forecast with mobile phone data

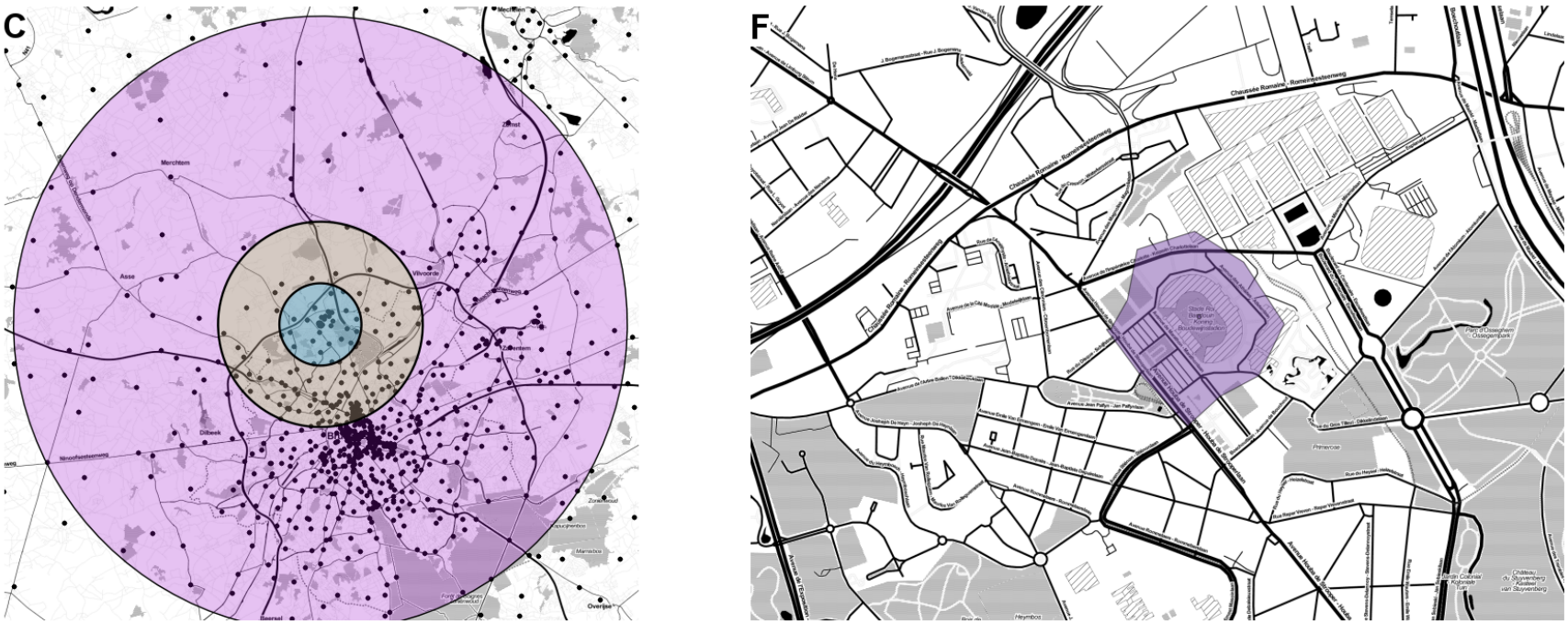
Dataset

- Call Detail Records: caller and callee IDs and cells, timestamp
- 5 – 6 March 2014
- Voice: 4.8×10^6 outgoing, 3.3×10^6 incoming
- Text: 19.9×10^6 outgoing, 18.7×10^6 incoming



Joint work with Vincent Blondel, submitted to Big Data Research (2014).

Measuring the fluxes of people

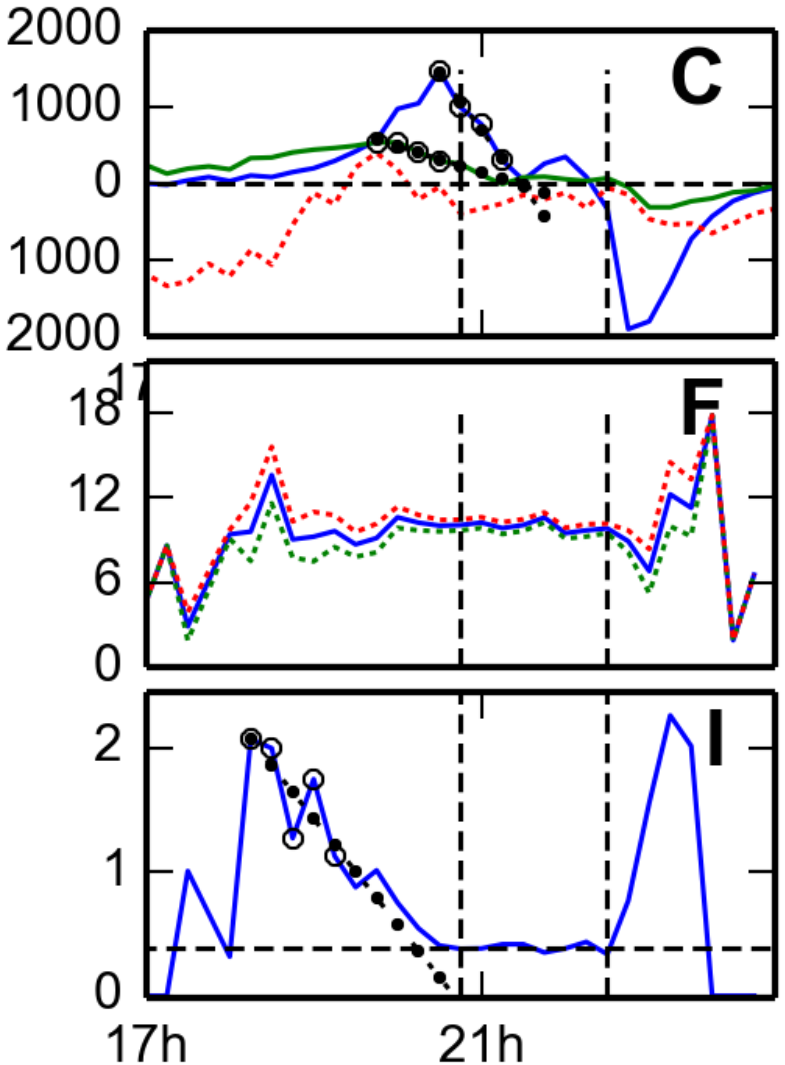


Concentric circles with radii 2, 5 and 15 km around the venues (left) and area within which the tweets were collected (right).

2 Methods

- Flux(r,t) = # people approaching - # people leaving
- StandardDeviation(distance to event | calling to event)

A forecasting of the zero fluxes is feasible



Subscribers fluxes (C), mean distance to the event $\langle d(t) \rangle$ of the text messages sent to the event (F) and standard deviation $\sigma_d(t)$ (I)

Perspectives

- More accurate models
- Use the social network
- Predictive calling behaviours

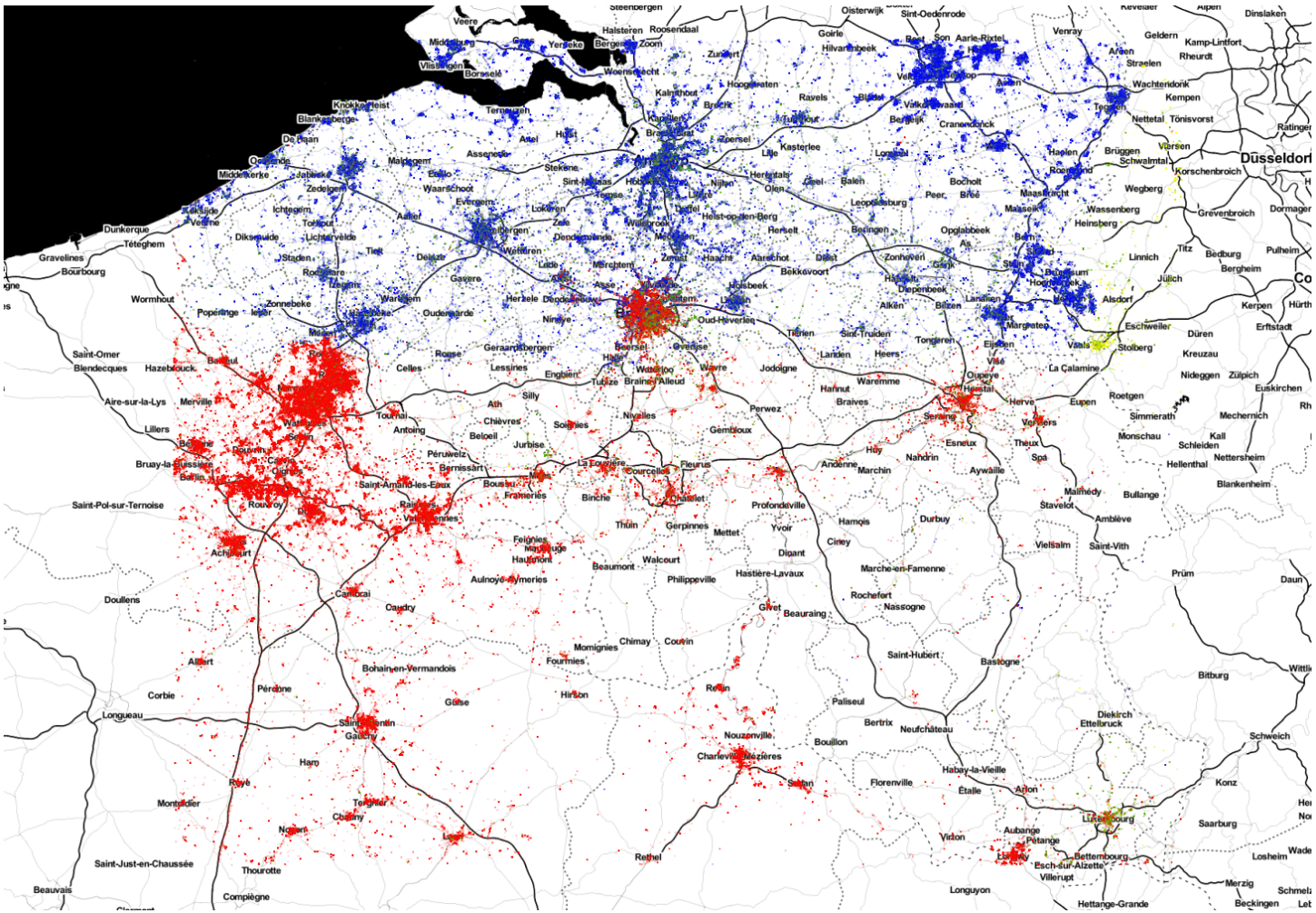
Twitter communities in Belgium: does space matter ?

Twitter

- Twitter Streaming API
- **Geotagged tweets** for Belgium
- $\sim 120,000$ users
- $\sim \mathbf{6.2 \cdot 10^6}$ **tweets**
- nodes=users having exchanged > 3 tweets, ties=reply-to.
- Resulting network has 8828 nodes and 13986 edges.

Work in progress joint with Vincent Blondel, Isabelle Thomas, Jean-Charles Delvenne.

Belgium is a bilingual country where French speaking people do not tweet a lot



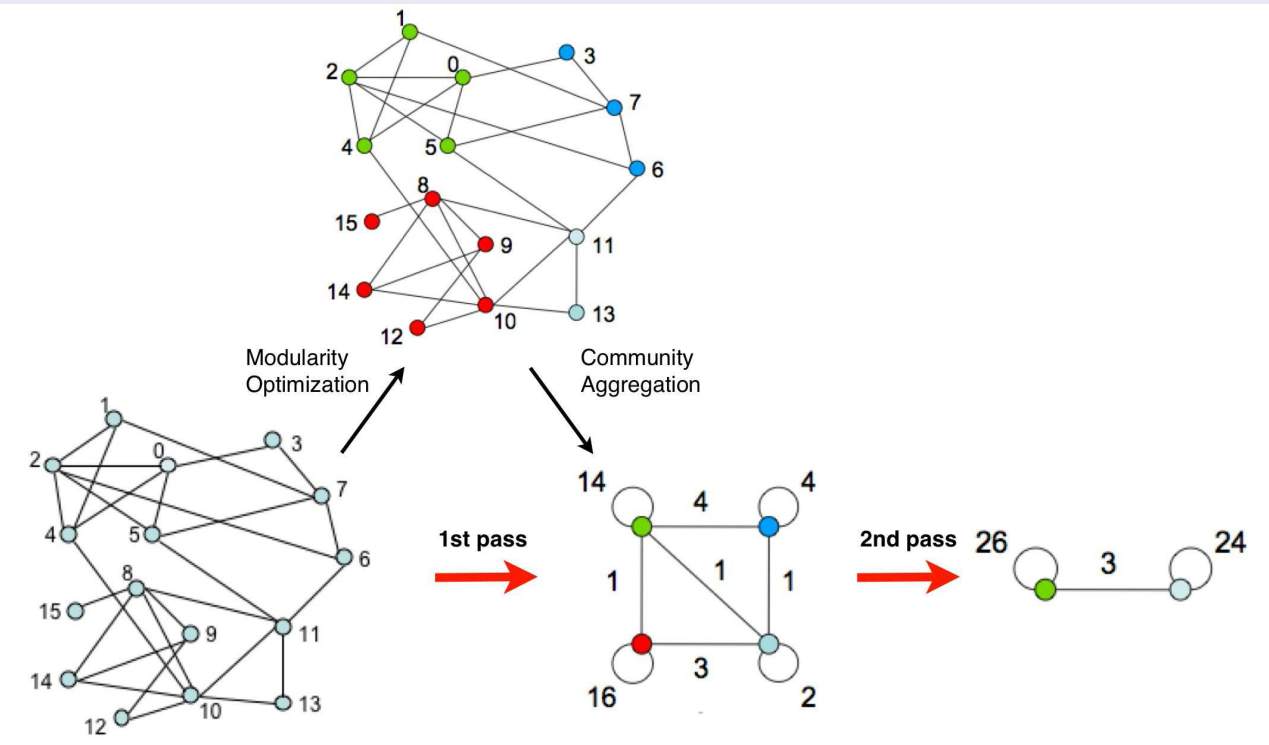
Language attributed to the tweet by Twitter

Community detection

Modularity optimization [Newman and Girvan, 2004]

$$Q = \frac{1}{2m} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(w_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Louvain method [Blondel et al, 2008]



Relevant scales

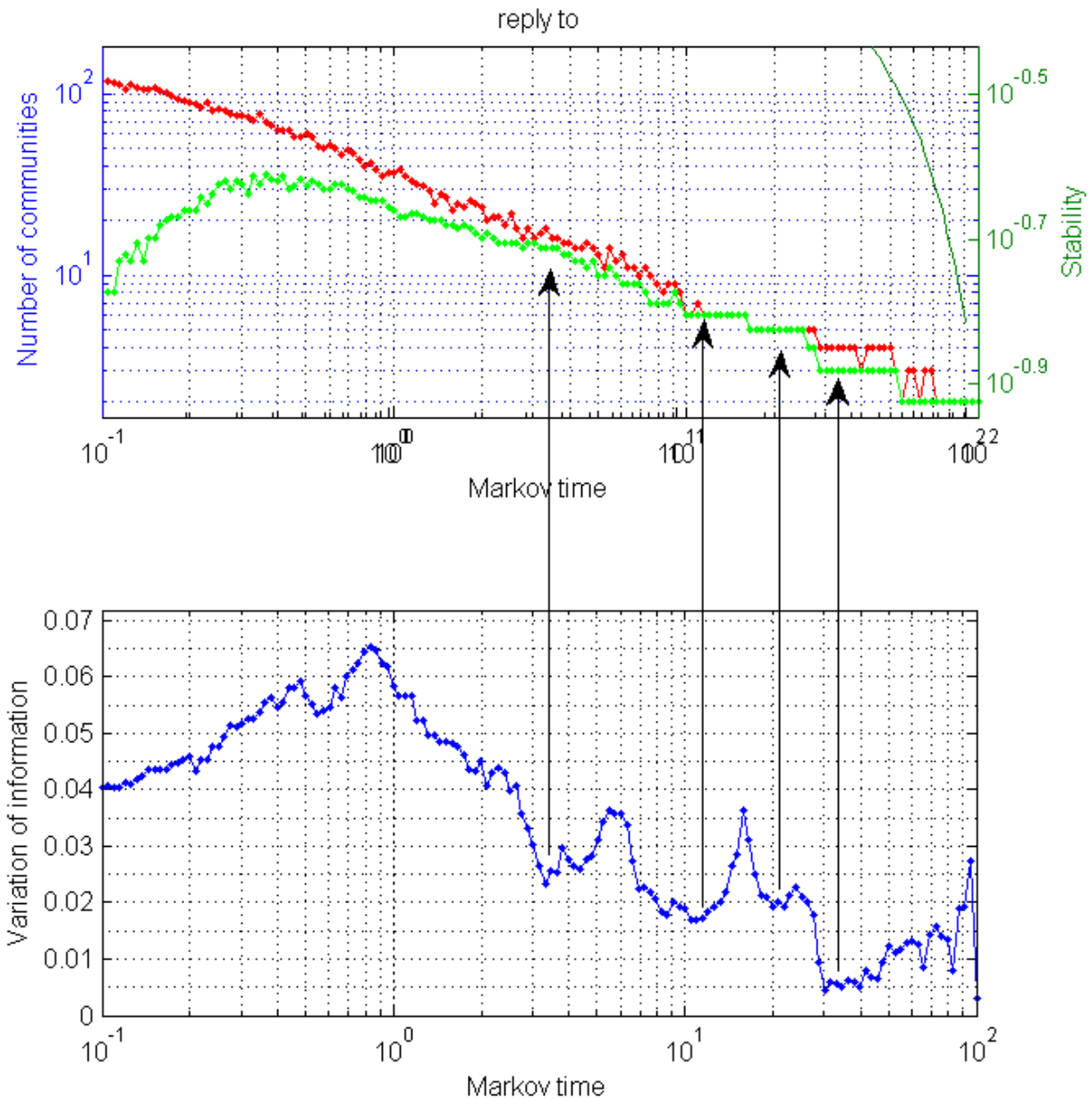
Relax the modularity

$$Q = \frac{1}{2\,m} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\mathbf{t} w_{ij} - \frac{k_i k_j}{2\,m} \right) \delta(c_i, c_j)$$

How to choose t ? [Delvenne et al, 2011]

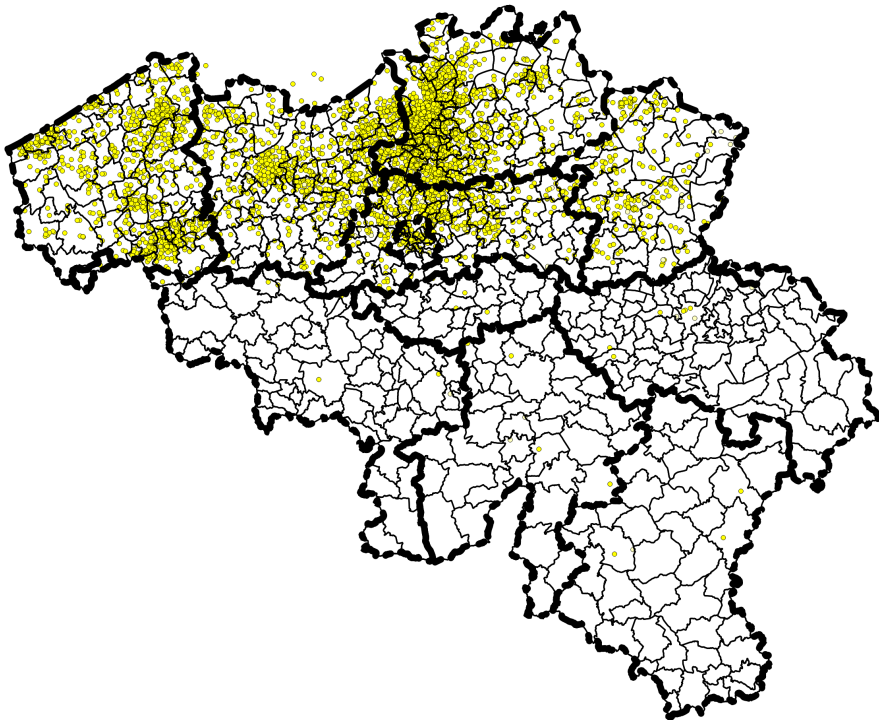
- Swipe t.
- For each t: compute the communities n times
- See if differ:
  - among the trials (low variation of information)
  - among the scales
- Relevant scales are those for which # of communities does not change

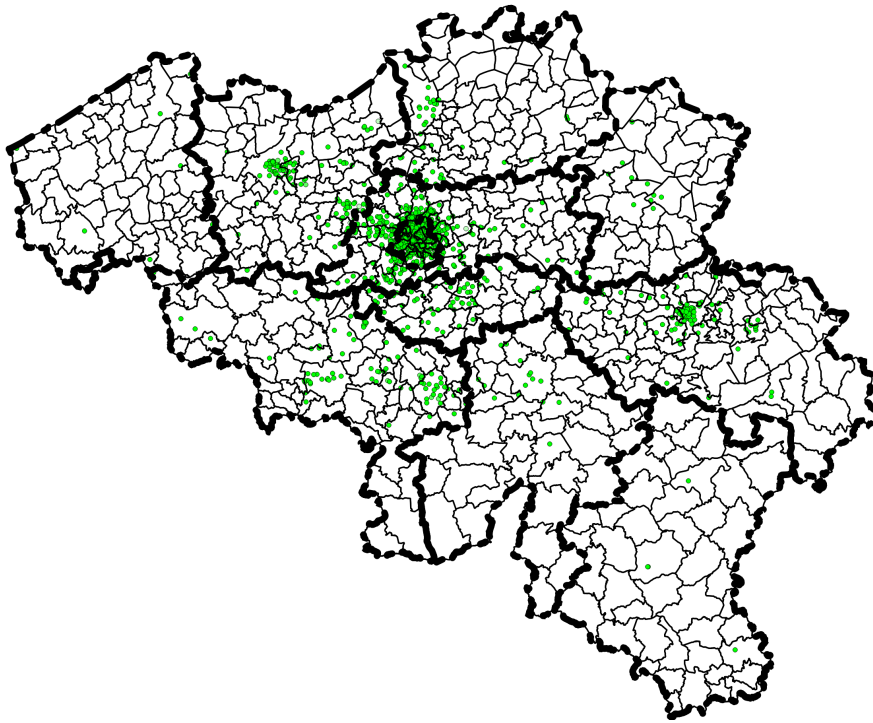
Four relevant scales for the reply-to network

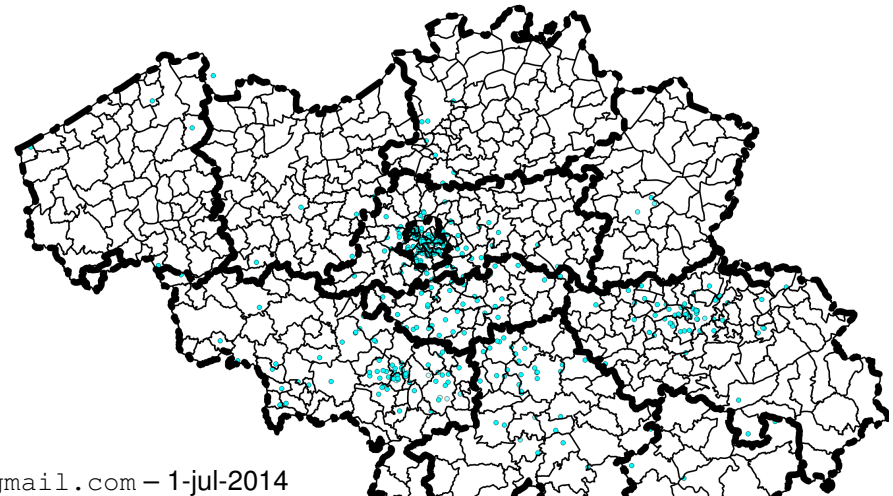


A cities network besides the language-based networks

t=35

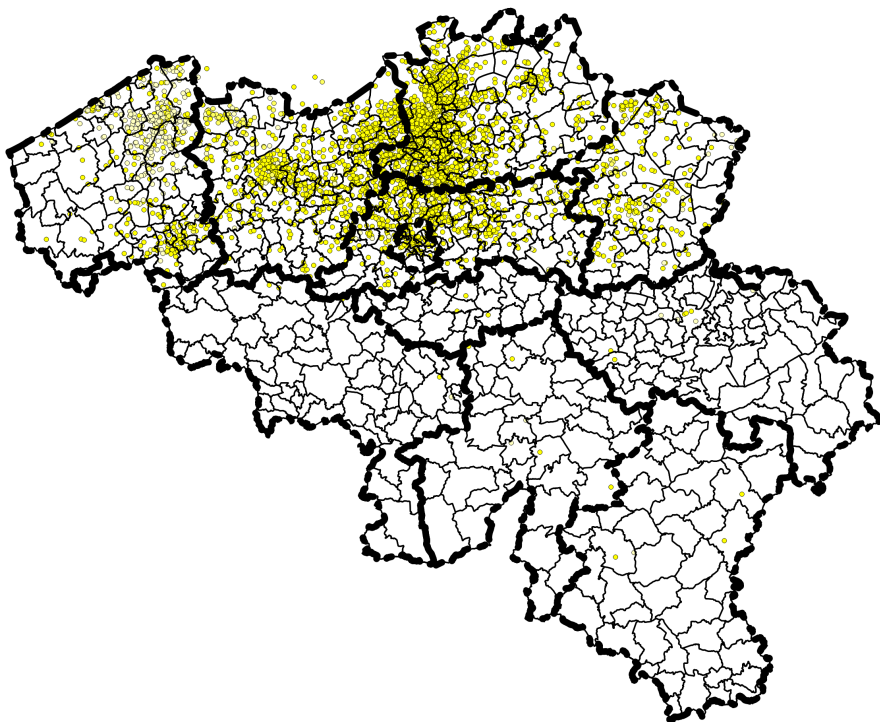


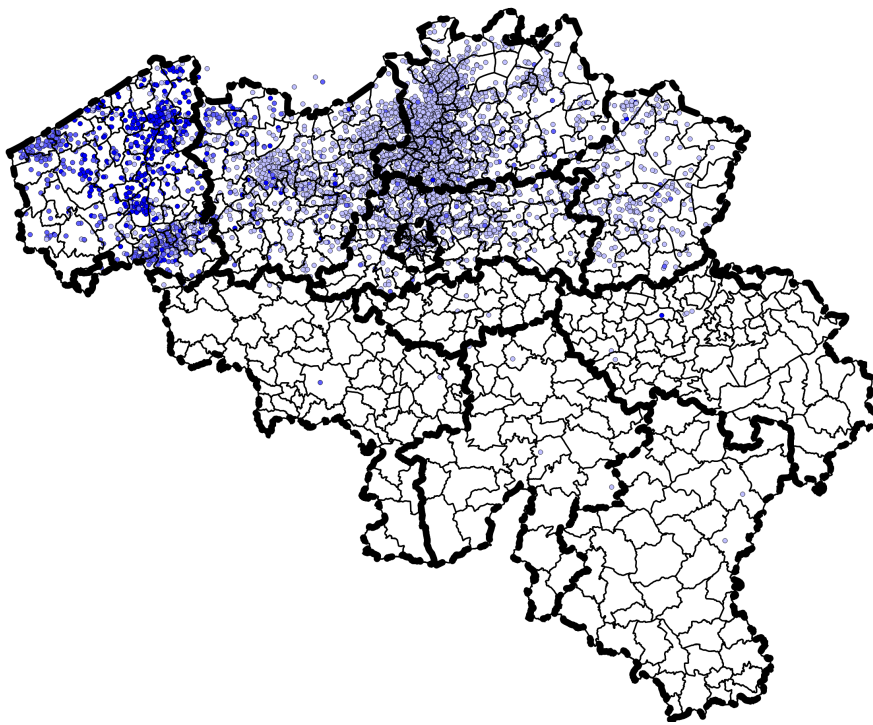




Flanders is structured around two poles

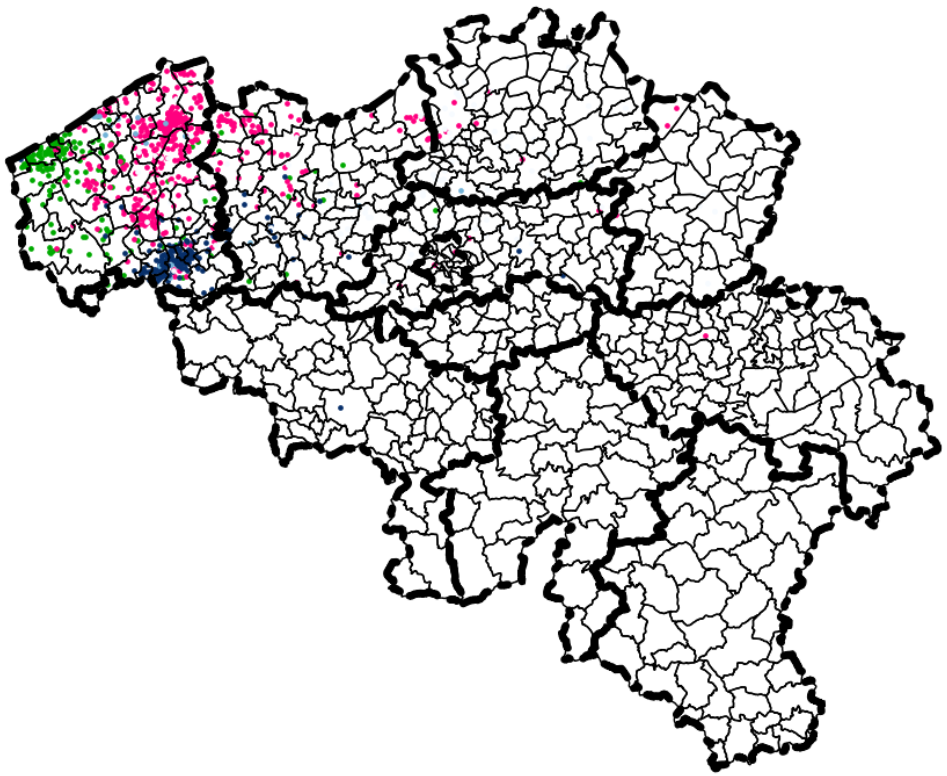
t=20





West-Flanders is structured around three cities

t=3.5



Perspectives

- Improve the network construction
- Address the drawbacks of modularity [Lancichinetti and Fortunato, 2011; Good et al, 2010; Lee and Cunningham, 2014, . . .]
  - Statistical significance ?
  - Overlapping communities ?
  - Local optimization ?
  - . . .
- By comparing with other techniques (eg: OSLOM [Lancichinetti et al, 2011,])

Conclusions

- Mobile phone data help to forecast the crowds
- Twitter communities in Belgium transcend linguistic communities.

Thank you

Christophe Cloquet
Université Catholique de Louvain (UCLouvain) – Belgium
post-doc until yesterday

christophe.cloquet@uclouvain.be → c.cloquet@gmail.com
@ibrux – linkedin.com/ccloquet

Joint works with Vincent Blondel (on crowd & twitter), Isabelle Thomas (on twitter) and Jean-Charles Delvenne (on twitter).