Locating the Political Spectrum of Amsterdam

Group 20

Abigail de la Rosa, Bojana Urumovska, Norman Knyazev, Salim Salmi and Athanasios Dritsas

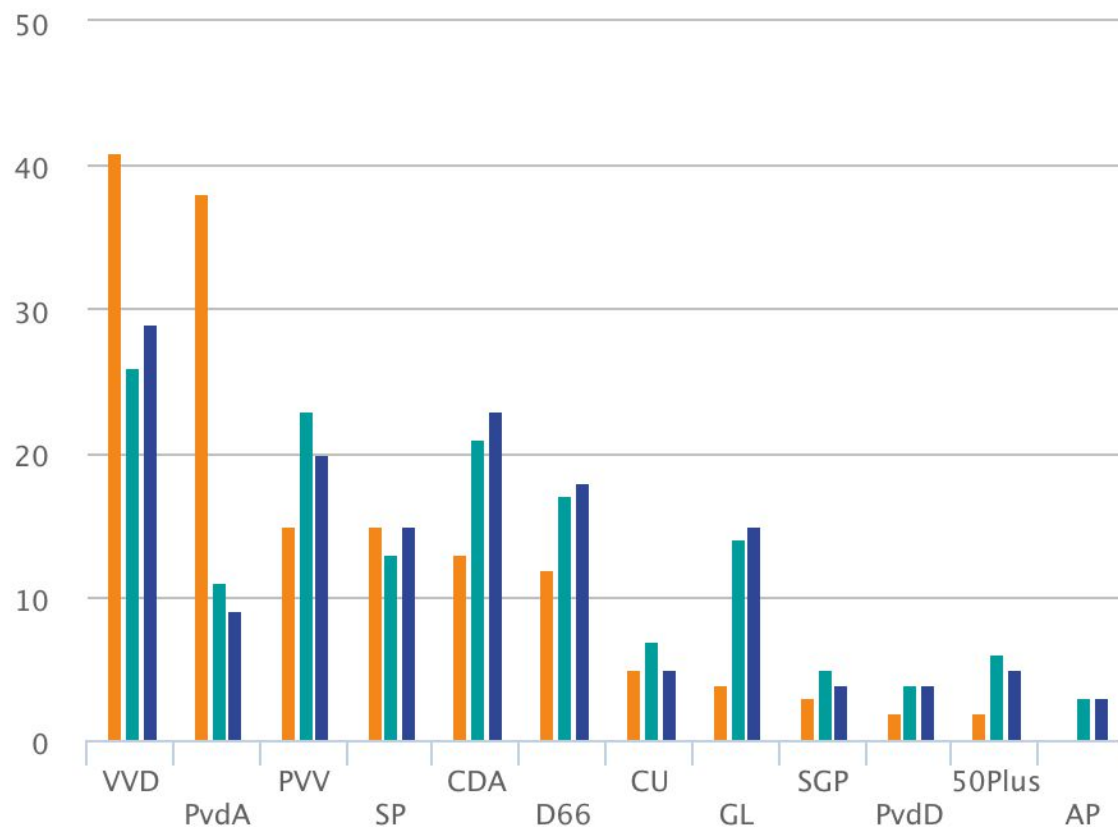Supervisor: Dr Achilleas Psyllidis

Weert Gilders

ZETELVERDELING LAATSTE TWEE PEILINGEN

Legend:
- Tweede Kamer (orange)
- 9 mrt '17 (teal)
- 14 mrt '17 (blue)

Parties: VVD, PvdA, PVV, SP, CDA, D66, CU, GL, SGP, PvdD, 50Plus, AP

Donald J. Trump ✓
@realDonaldTrump

"@susanbirchfiel1: Nothing you say will change my mind! Woman for @realDonaldTrump twitter.com/Carolde/status ..."

> **Carol** @Carolde
> Dear @CNN @MSNBC @FoxNews #TrumpSupporters are sick of the #Trump bashing RT and or add a comment

RETWEETS	LIKES
5,759	13,813

4:57 AM - 31 Mar 2016

↩ 2.5K   ⟲ 5.8K   ♥ 14K

What do we offer?

- Untapped social media data source
- Localised
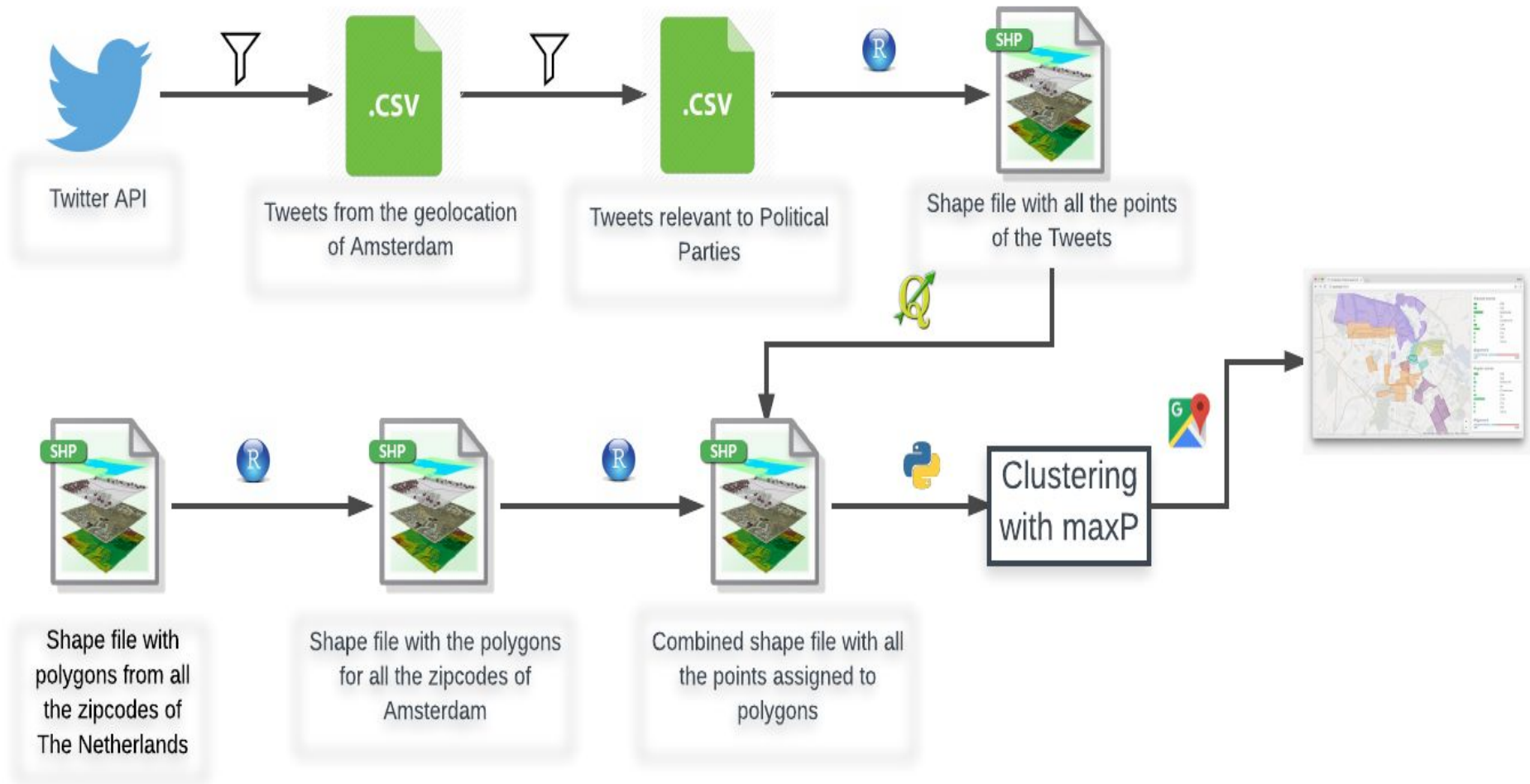- Up to date information
- Aggregate individuals into regions

Who is it for?

- Politicians
- Marketing purposes
- Individuals

How does it work?

Twitter Developer Documentation

Products & Services

POST statuses/filter

Returns public statuses that match one or more filter predicates. Multiple parameters may be specified which allows most clients to use a single connection to the Streaming API. Both GET and POST requests are supported, but GET requests with too many parameters may cause the request to be rejected for excessive URL length. Use a POST request to avoid long URLs.

The track, follow, and locations fields should be considered to be combined with an OR operator. `track=foo&follow=1234` returns Tweets matching "foo" OR created by user 1234.

The default access level allows up to 400 track keywords, 5,000 follow userids and 25 0.1-360 degree location boxes. If you need elevated access to the Streaming API, you can contact Gnip.

Resource URL

https://stream.twitter.com/1.1/statuses/filter.json

Groenlinks

Arrived at AFAS Live for the meetup of #GroenLinks! #Klaver #voteforchange
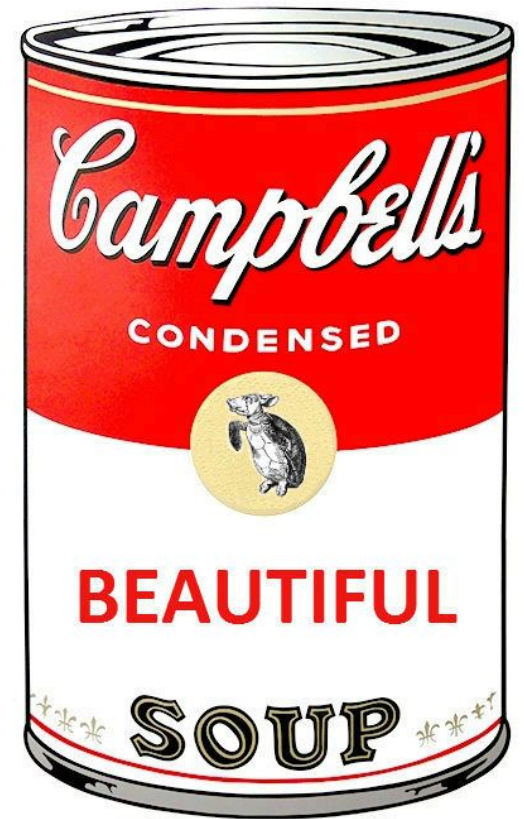
Techniques

Crawling and Scraping

- 10 Domains
- 1000 Pages/Domain
- Tf-idf

PvdA: Progressive, Socialisten

PVV: Nikaab, Sharia

50plus: Pensioen, Ouderen

PvdA?

Fabrice
@GrimlocK38

Follow

On Air 🎧 Progressive Pick 👌
#asotfestival #trance #trancefamily
#arminvanbuuren #dj #artist… instagram.com
/p/BRtgw51ADX1/

LIKES
2

12:42 PM - 16 Mar 2017 from Amsterdam, The Netherlands

2

Hans U. de Jong
@hansudejong

Follow

Eindelijk te koop, het eerste duurzaam geproduceerde commerciële Broodbier in Nederland @...
instagram.com/p/BRVpFxoAOQo/

🌐 Translate from Dutch

2:13 PM - 7 Mar 2017 from Waterland, Nederland

Keyword:

Duurzaam

English:

Sustainable

Techniques - Clustering

Regionalization - Spatial clustering

- Goal : Clustering of zipcode areas into larger regions based on their political alignment
- Constraint : Areas being in the same cluster need to be geographical neighbors
- Standard clustering algorithms cannot cope with this constraint
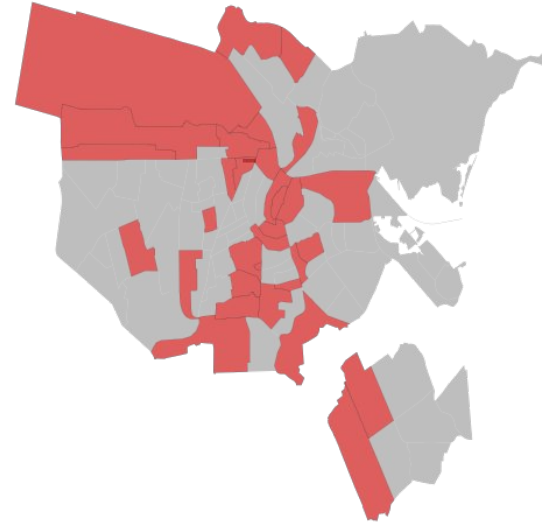- Regionalization methods are required

Max-p algorithm

- Number of regions not known beforehand
- Minimum threshold for a specific variable per each region
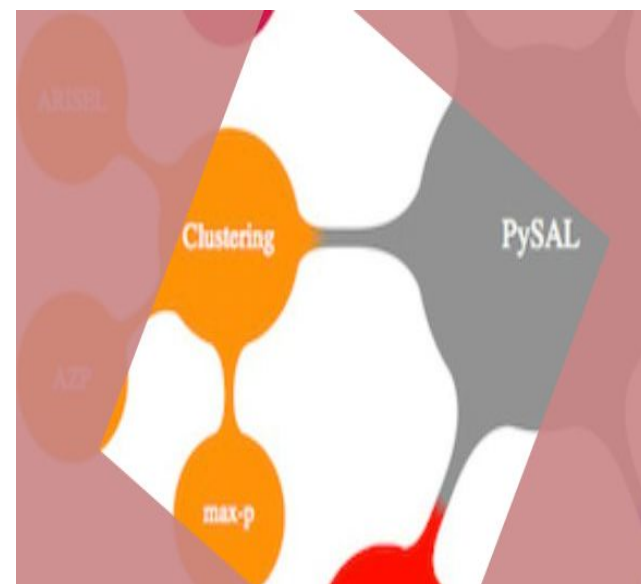
Techniques - Clustering

Spatial weight matrix

- 3-NN Nearest Neighbors algorithm
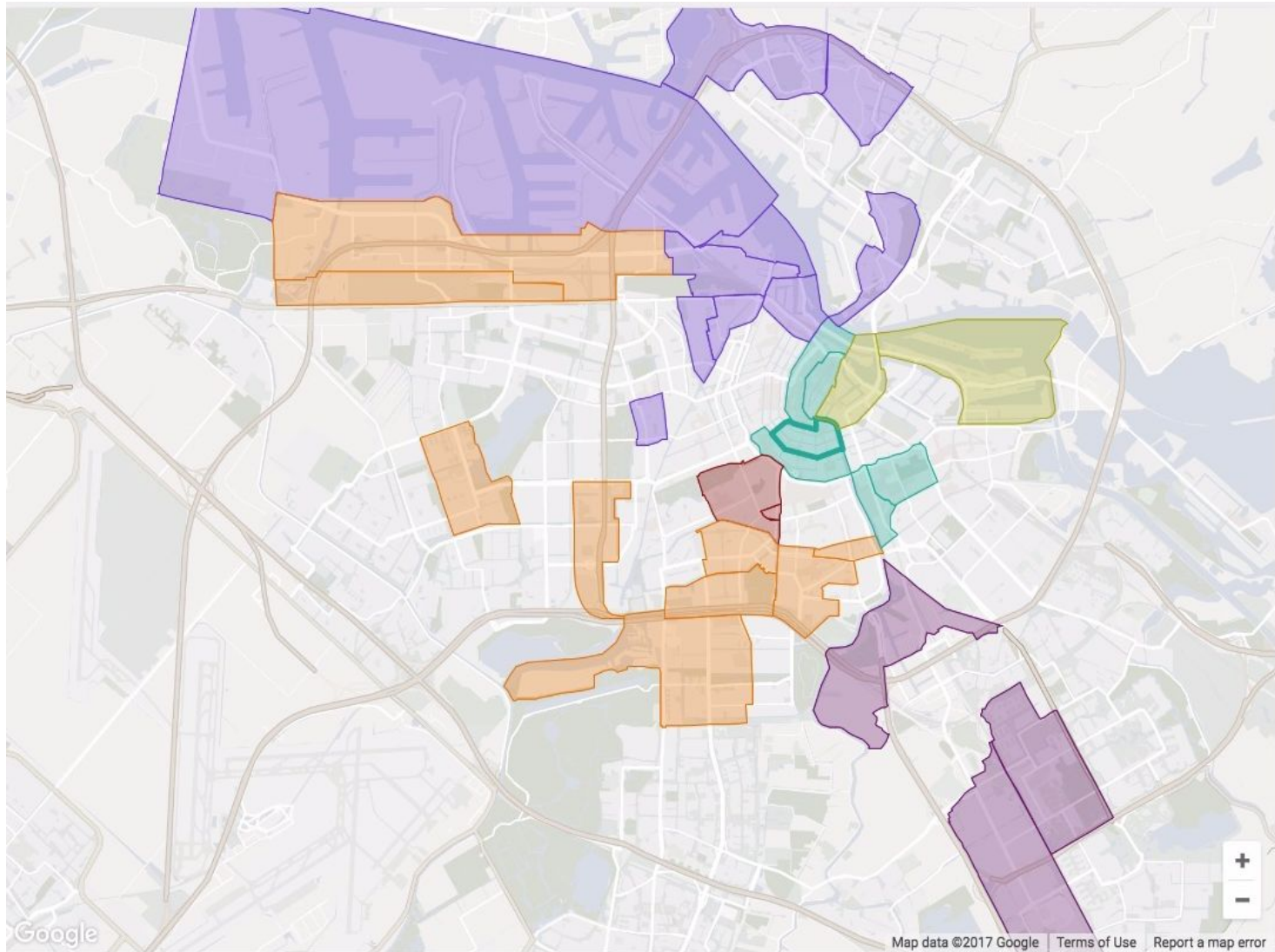- Queen and Rook contiguity approaches lead to islands

Political alignment attribute

- Weighted sum of the percentage of tweets per party
- Weights represent position of a party in the political spectrum
- Left:0, Centre-left:0.25, Centre:0.5 , Centre-right:0.75, Right:1

Zipcode	VVD	GL	D66	Political Alignment
1013	3	3	4	(3*0.75+3*0+4*0.5)/10 = 0.425

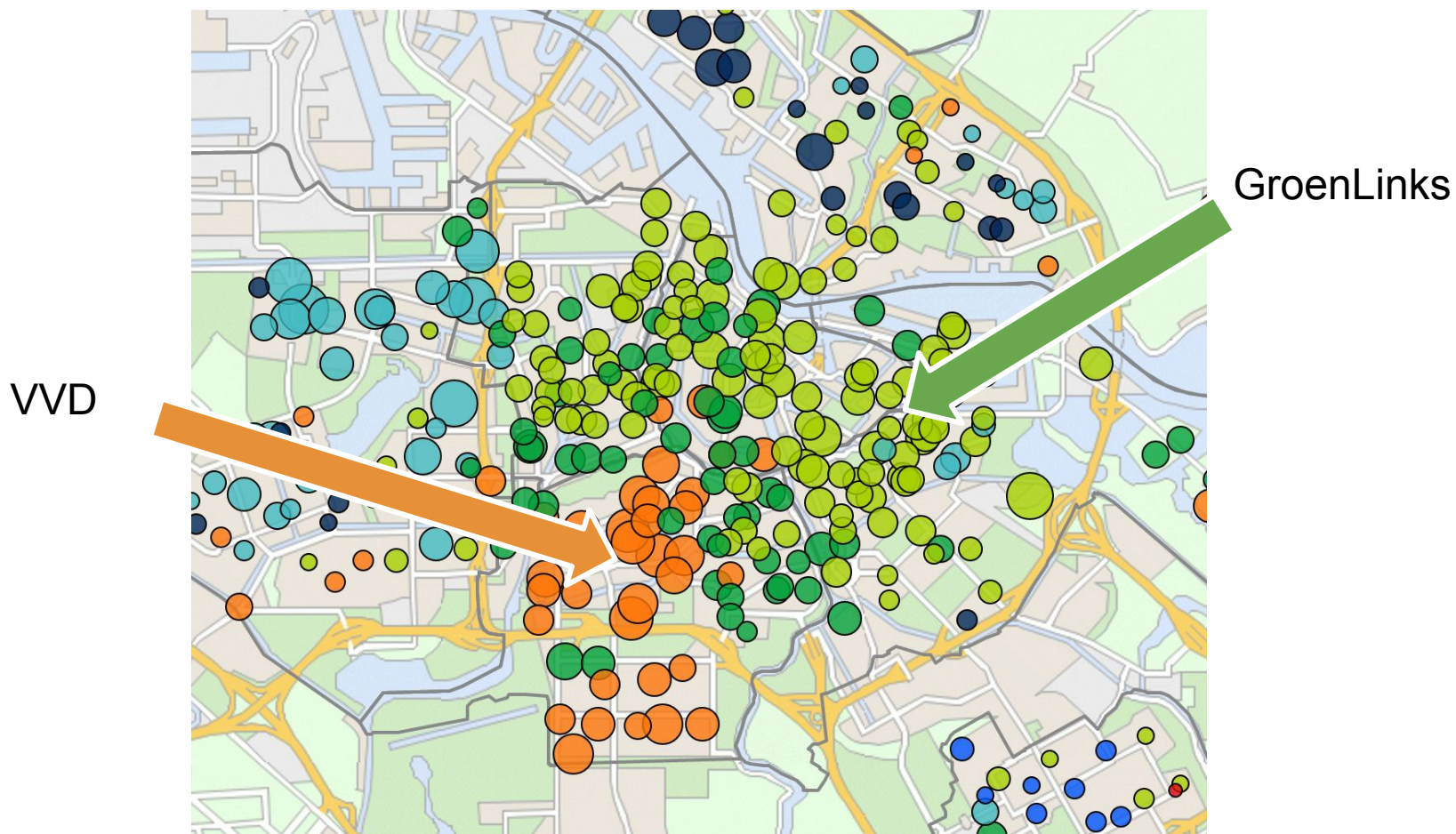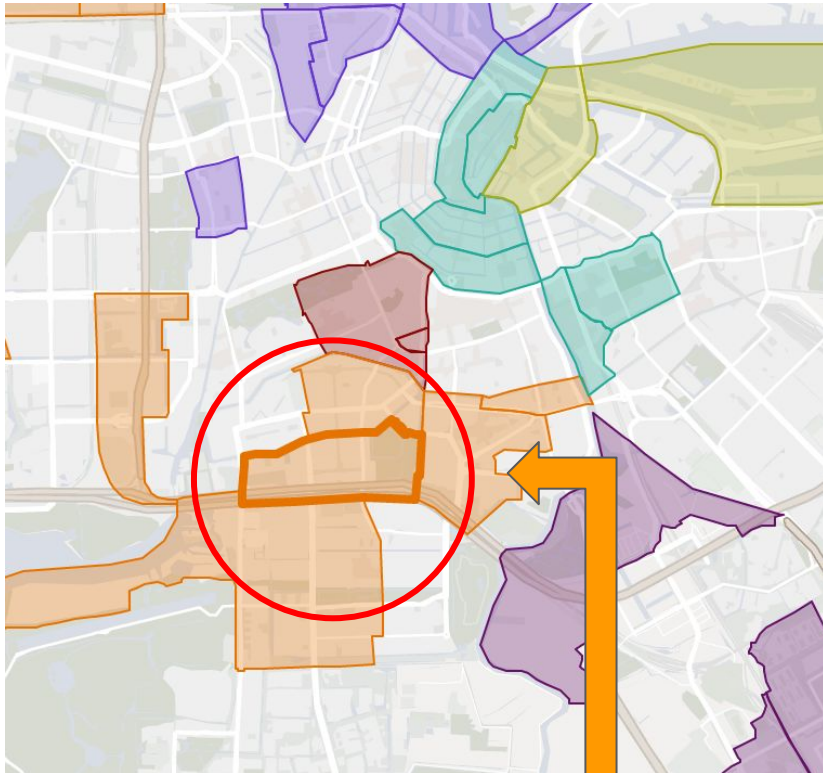# Techniques - Clustering

Techniques - Visualisation
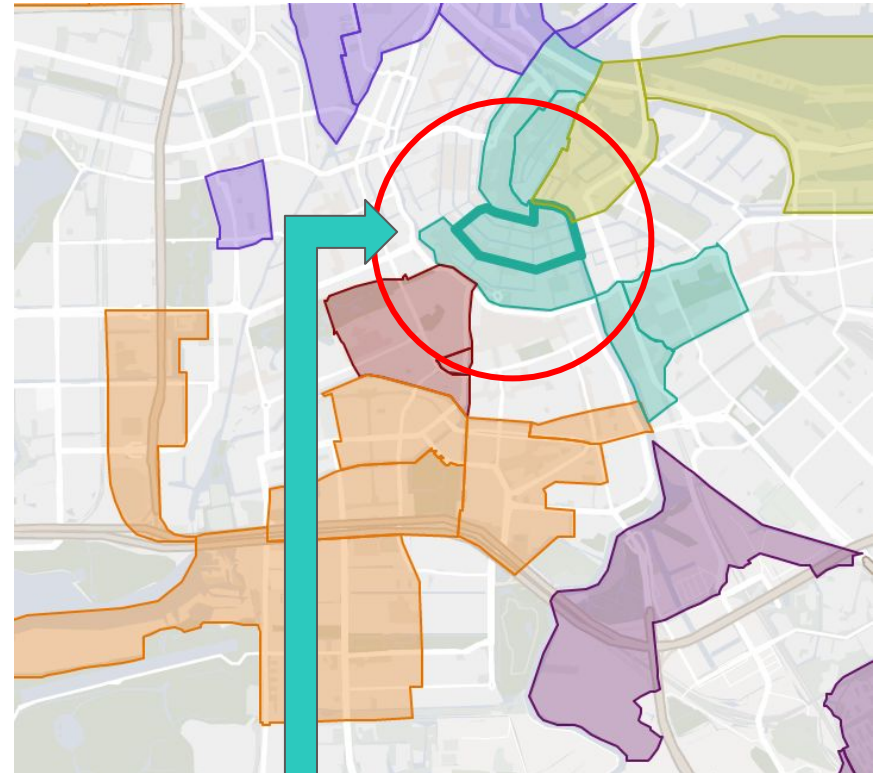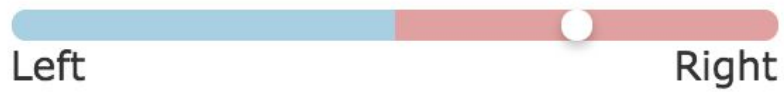
Google maps interactive map

Demo
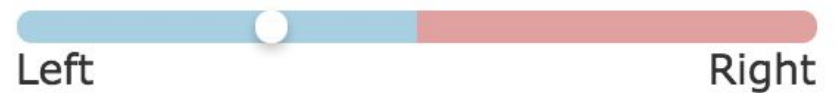
Evaluation
Results per polling station

GroenLinks

VVD

Evaluation
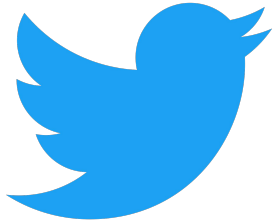
Recap

Minimize:

$$Z = \left(-\sum_{k=1}^{n}\sum_{i=1}^{n} x_i^{k0}\right) * 10^h + \sum_{i}\sum_{j|j>i} d_{ij} t_{ij}$$

Subject to:

$$\sum_{i=1}^{n} x_i^{k0} \leq 1 \qquad \forall k = 1, \cdots, n$$

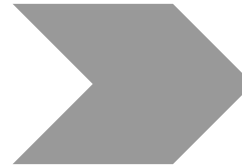$$\sum_{k=1}^{n}\sum_{c=0}^{q} x_i^{kc} = 1 \qquad \forall i = 1, \cdots, n$$

$$x_i^{kc} \leq \sum_{j \in N_i} x_j^{k(c-1)} \qquad \forall i = 1, \cdots, n; \forall k = 1, \cdots, n; \forall c = 1, \cdots, q$$

$$\sum_{i=1}^{n}\sum_{c=0}^{q} x_i^{kc} l_i \geq threshold * \sum_{i=1}^{n} x_i^{k0} \qquad \forall k = 1, \cdots, n$$

$$t_{ij} \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} - 1 \qquad \forall i, j = 1, \cdots, n | i < j; \forall k = 1, \cdots, n$$

$$x_i^{kc} \in \{0,1\} \qquad \forall i = 1, \cdots, n; \forall k = 1, \cdots, n; \forall c = 0, \cdots, q$$

$$t_{ij} \in \{0,1\} \qquad \forall i, j = 1, \cdots, n | i < j$$

References

1. Duque, J. C., Anselin, L., & Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science*, *52*(3), 397-419.
2. Rey, S.J. & Anselin, L. (2010) PySAL: a python library of spatial analytical methods. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (eds M.Fischer & A.Getis), pp. 175–193. Springer, Berlin.