Social Media, Networks and Geographical Data

Telmo Menezes and Camille Roth
{menezes, roth}@cmb.hu-berlin.de

Centre Marc Bloch - Computation Social Science Team

Studying Border Regions in the Post-Soviet Space

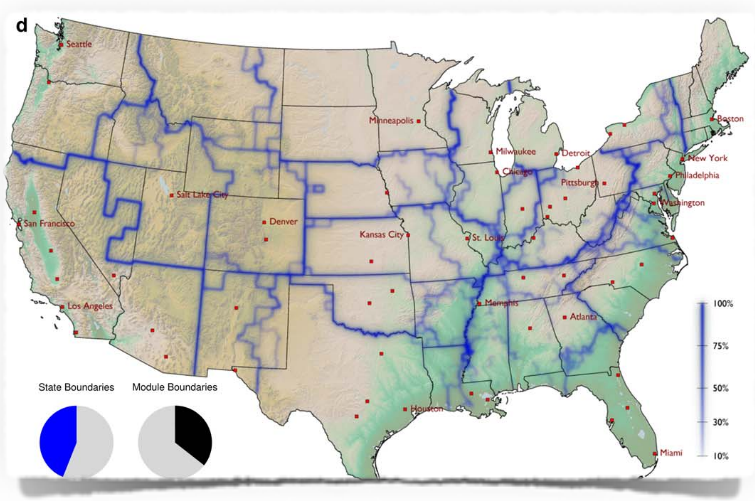
Different Methods, Scales and Areas

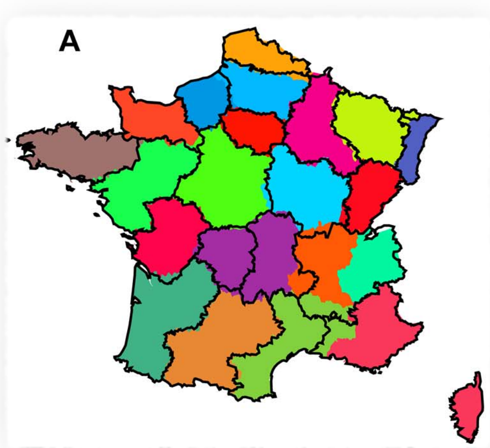
26-28 November 2019, Tbilisi State University, Georgia

Geographic “Communities”

Banknotes

- Thiemann, Theis, Grady, Brune, Brockmann, "**The Structure of Borders in a Small World**", *PLoS One* 5(11):e15422, **2010**





Cell phone data

- Ratti, Sobolevsky, Calabrese, Andris, Reades, Martino, Claxton, Strogatz, "**Redrawing the Map of Great Britain from a Network of Human Interactions**", *PLoS One*, 5(12):e14248, **2010**

- Calabrese, Dahlem, Gerber, Paul, Chen, Rowland, Rath, Ratti, **"The Connected States of America: Quantifying Social Radii of Influence",** *IEEE 3rd Intl Conf on Social Computing (SocialCom)*, pp. 223-230, **2011**

- Sobolevsky, Szell, Campari, Couronné, Smoreda, Ratti, **"Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries”,** *PLoS ONE* 8(12): e81707, **2013**

Check-in data

- Cranshaw, Schwartz, Hong, Sadeh, **"The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City"**, *ICWSM,* 58-65, **2012**

- Liu, Sui, Kang, Gao, **"Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data",** *PLoS One*, 9(1):e86026, **2014**

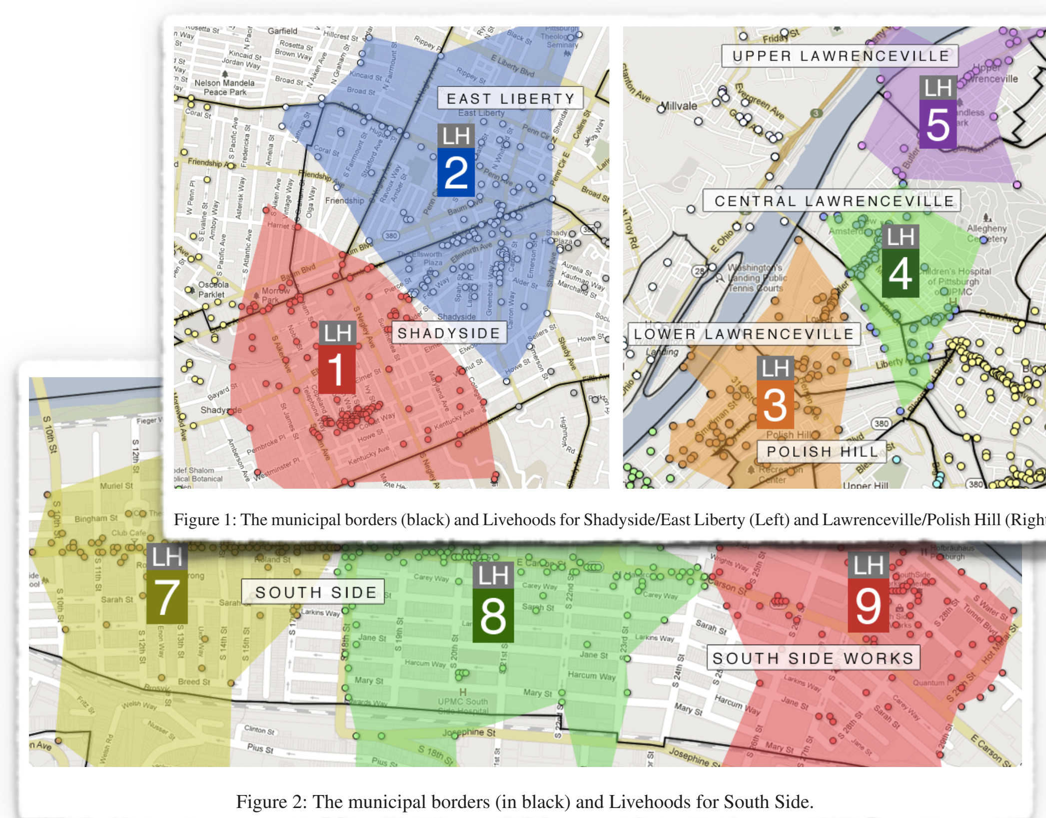
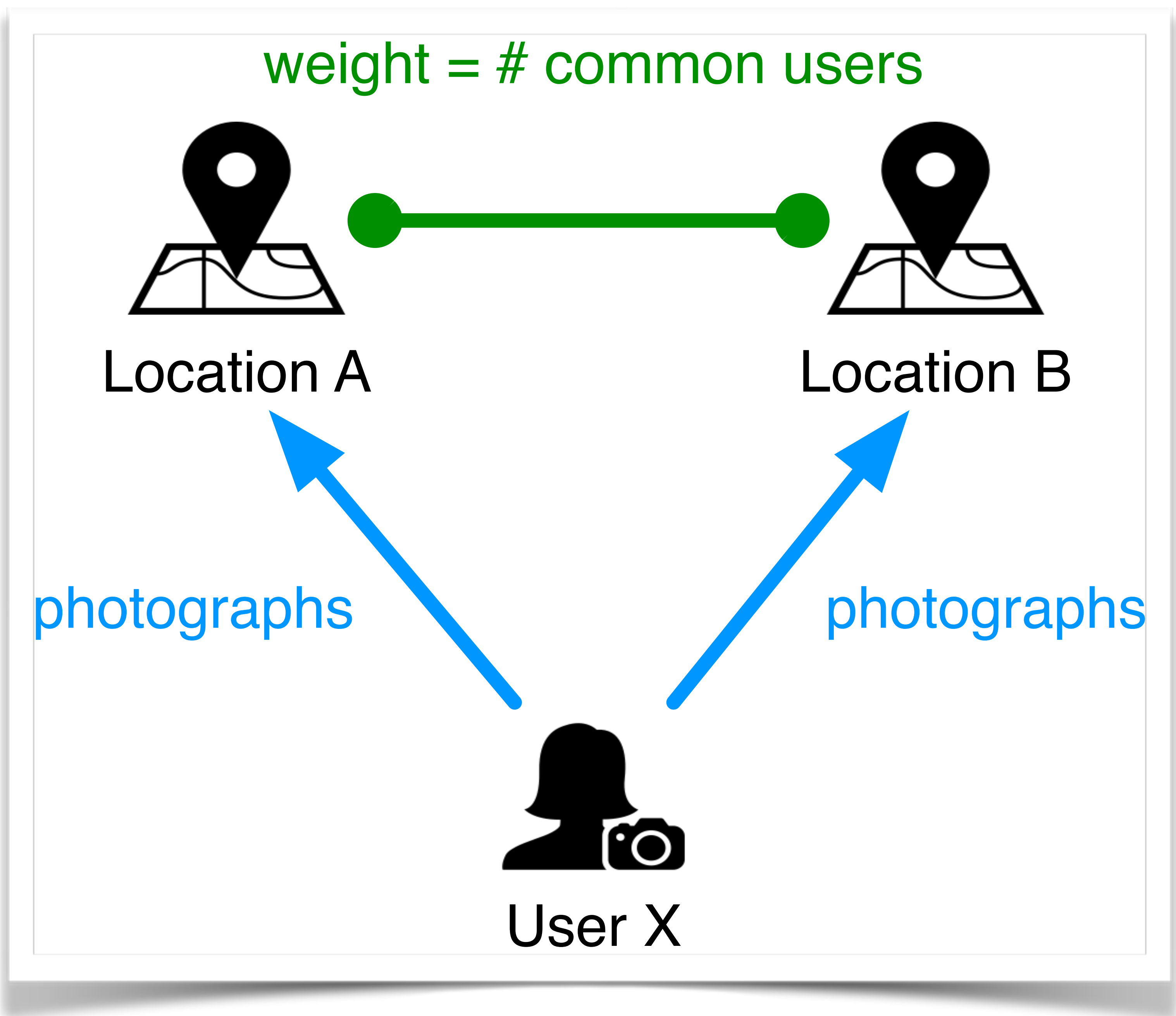


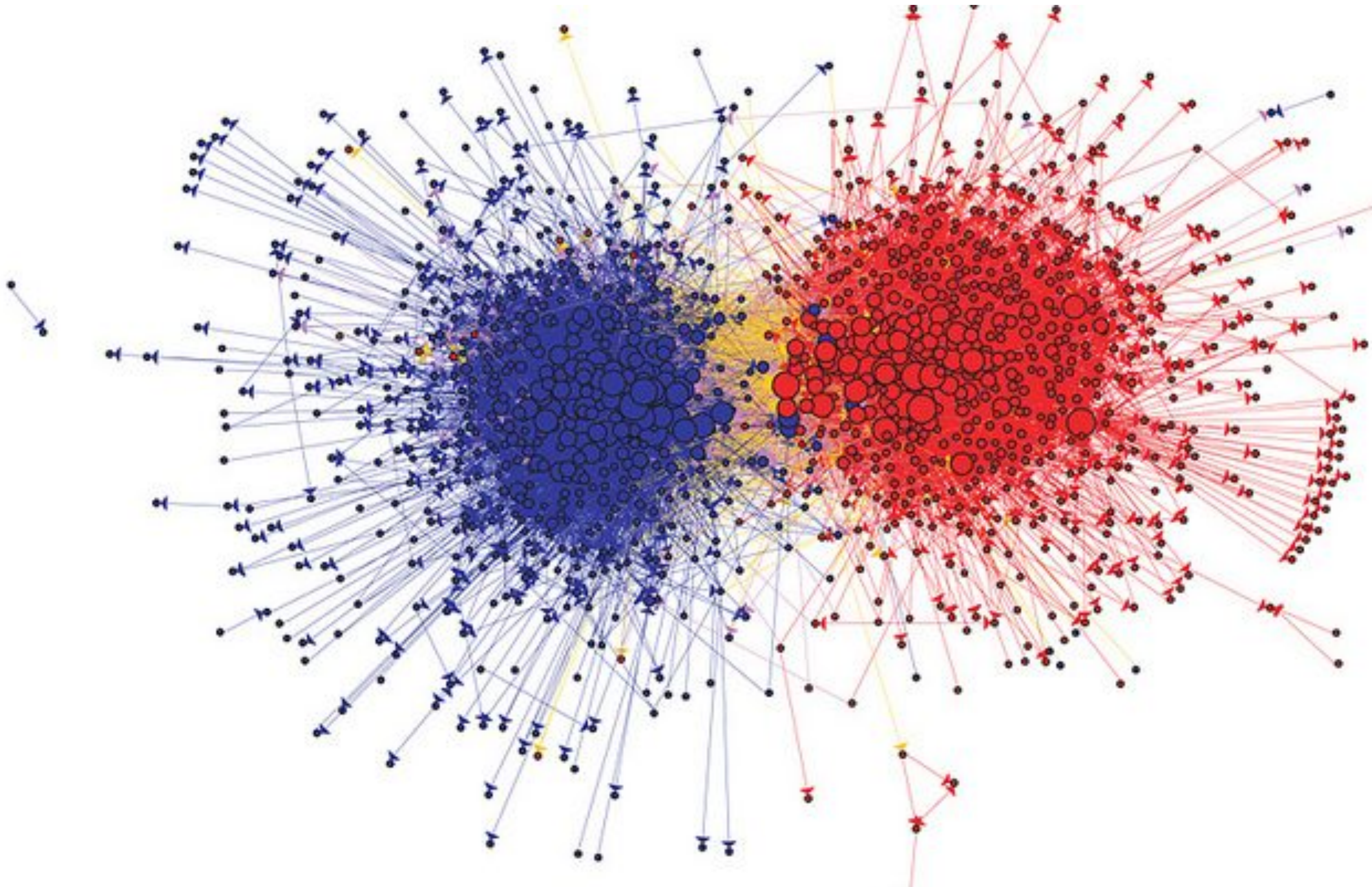
Figure 1: The municipal borders (black) and Livehoods for Shadyside/East Liberty (Left) and Lawrenceville/Polish Hill (Right).

Figure 2: The municipal borders (in black) and Livehoods for South Side.

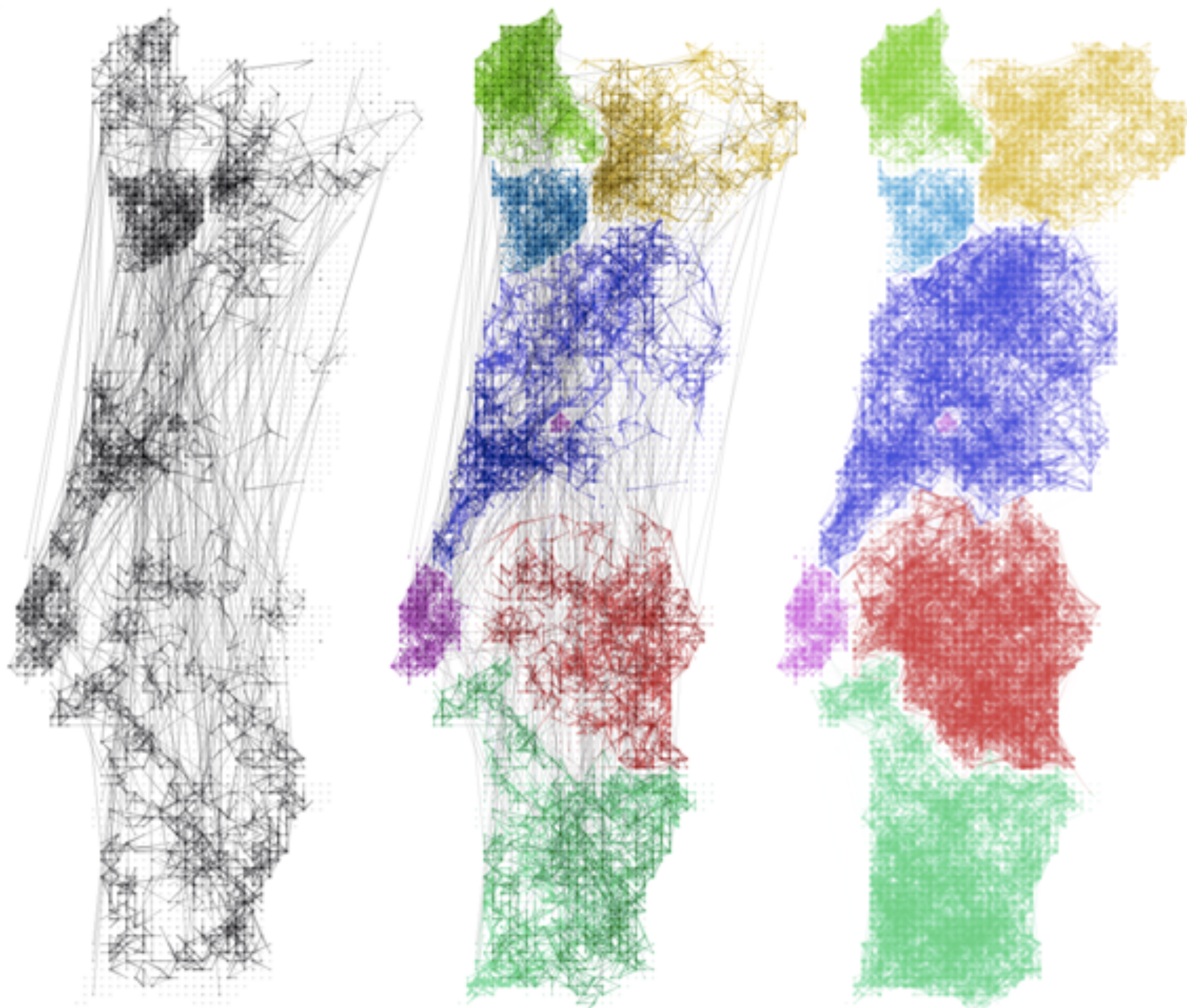
Instagram Colocation Network



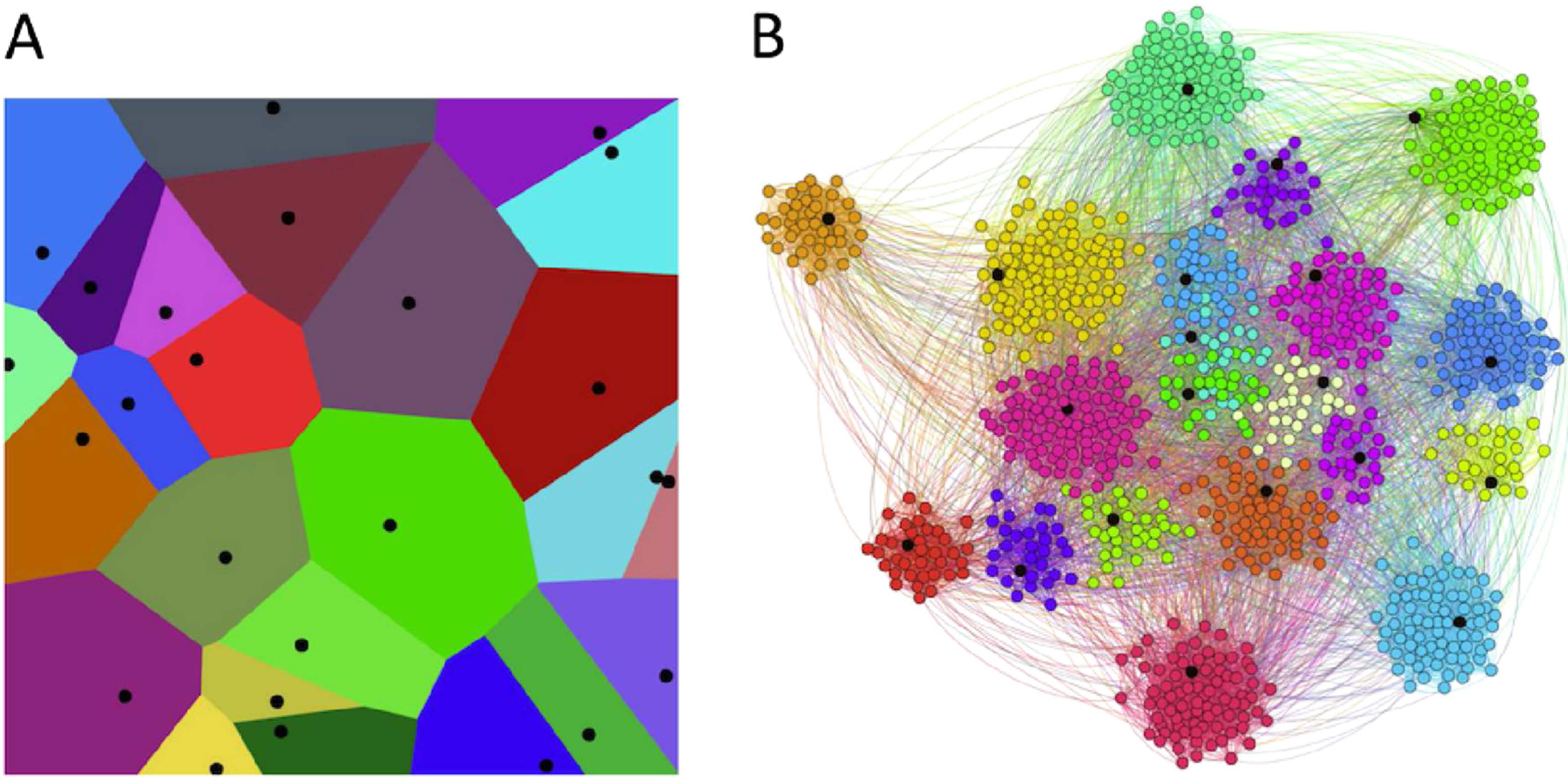
Polarization in political discussion (blog network)

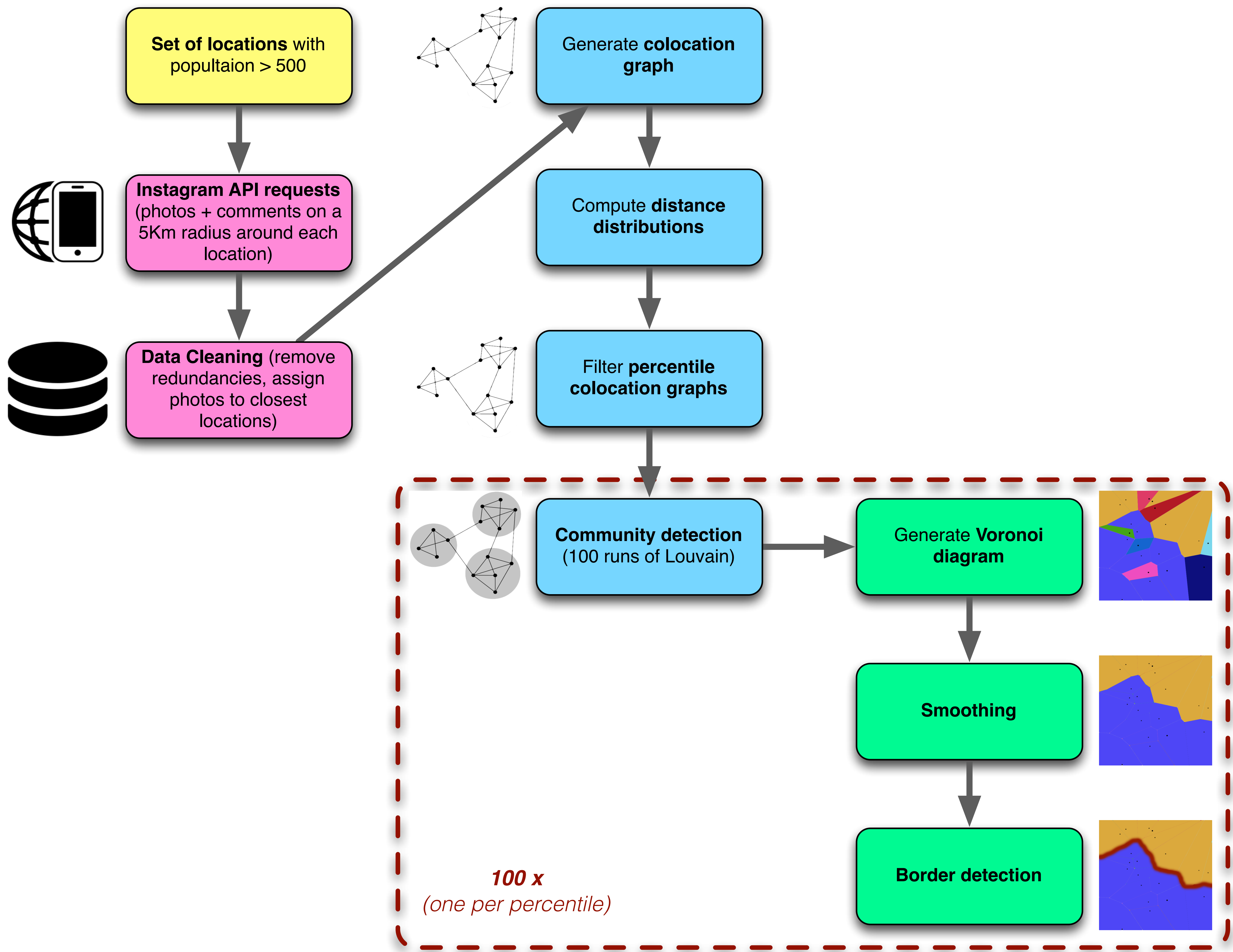


Geographically embedded network

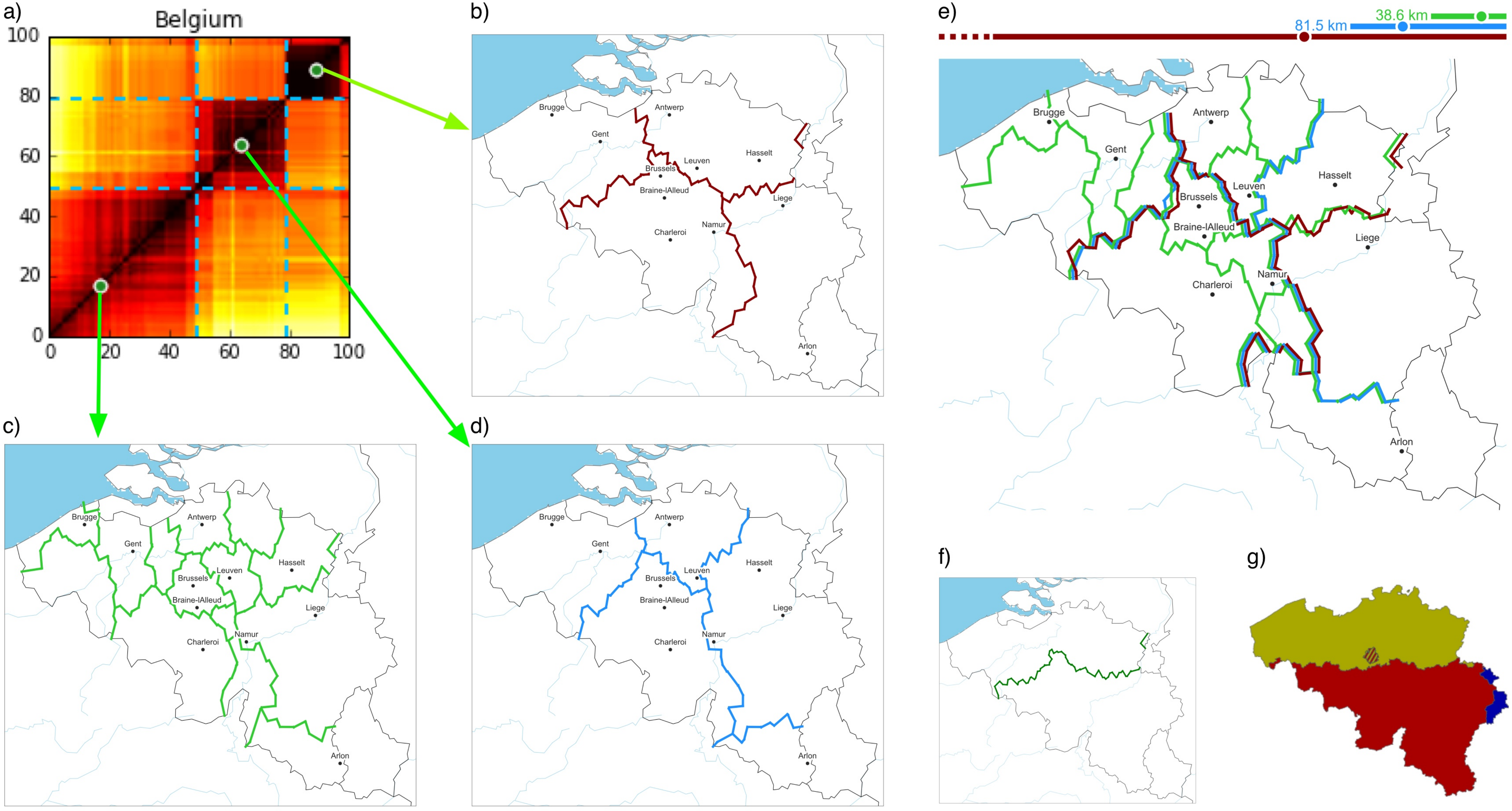


Voronoi diagrams and network communities





Set of locations with populaion > 500
Instagram API requests (photos + comments on a 5Km radius around each location)
Data Cleaning (remove redundancies, assign photos to closest locations)
Generate colocation graph
Compute distance distributions
Filter percentile colocation graphs
Community detection (100 runs of Louvain)
Generate Voronoi diagram
Smoothing
Border detection
100 x
(one per percentile)



a)
Belgium
b)
c)
d)
e)
f)
g)
81.5 km
38.6 km
Brugge
Gent
Antwerp
Hasselt
Leuven
Brussels
Braine-lAlleud
Liege
Namur
Charleroi
Arlon

Measuring Partition Similarity

Let us define a function $\mu^P(i,j)$ that takes the value 1 if both i and j belong to the same subset of a partition P, 0 otherwise:

$$\mu^P(i,j) = \begin{cases} 1, & \text{if } \exists X \in P \text{ such that } i,j \in X \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

We can then define the similarity of P_s and $P_{s'}$ as the ratio between the number of pairs of locations in V that have the same value of μ for both P_s and $P_{s'}$ (i.e. they are classified similarly at scales s and s'), and the total number of possible location pairs:

$$\delta(V, P_s, P_{s'}) = \frac{|\{(i,j) \in V^2, i \neq j, \mu^{P_s}(i,j) = \mu^{P_{s'}}(i,j)\}|}{\binom{|V|}{2}} \tag{3}$$

Intervals of Similar Scales

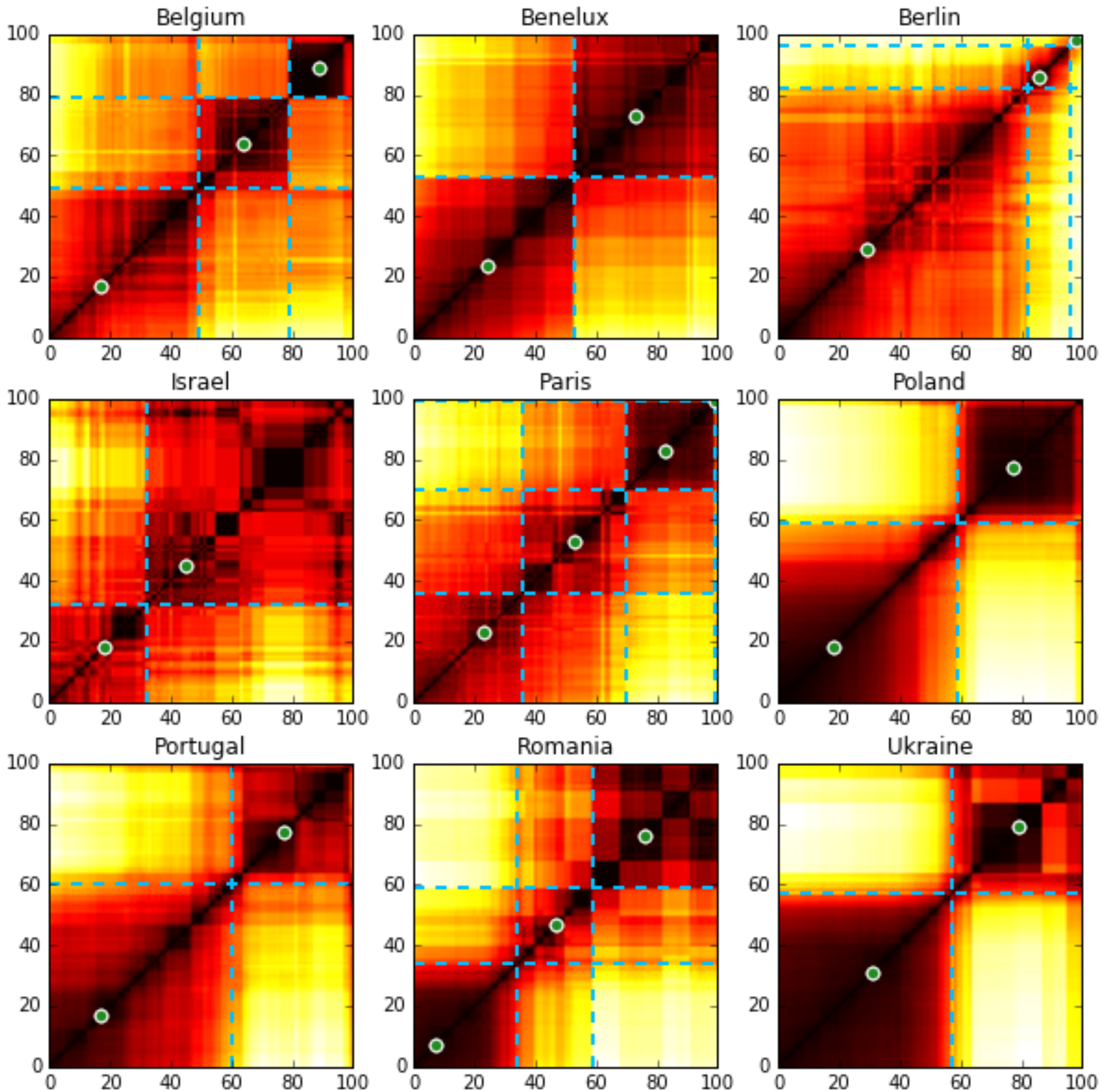
To identify the breakpoints in partition similarity we introduce another metric, somewhat similar to the concept of modularity in graphs – albeit even simpler. This metric measures *interval separation*, given a set of breakpoints $B = \{b_0, ..., b_n\}$. Let us also consider the set of intervals defined by these breakpoints: $\mathcal{I}(B) = \{]0..b_0],]b_0, b_1], ...,]b_n, 100]\}$. The interval separation for a given B can thus be defined as:

$$\sigma_B = \frac{\sum_{I \in \mathcal{I}(B)} |I| \cdot \sum_{s,s' \in I} \delta(V, P_s, P_{s'})}{\max\limits_{b \in B \setminus \{b_0\}} \delta(V, P_{b-1}, P_b)} \tag{4}$$

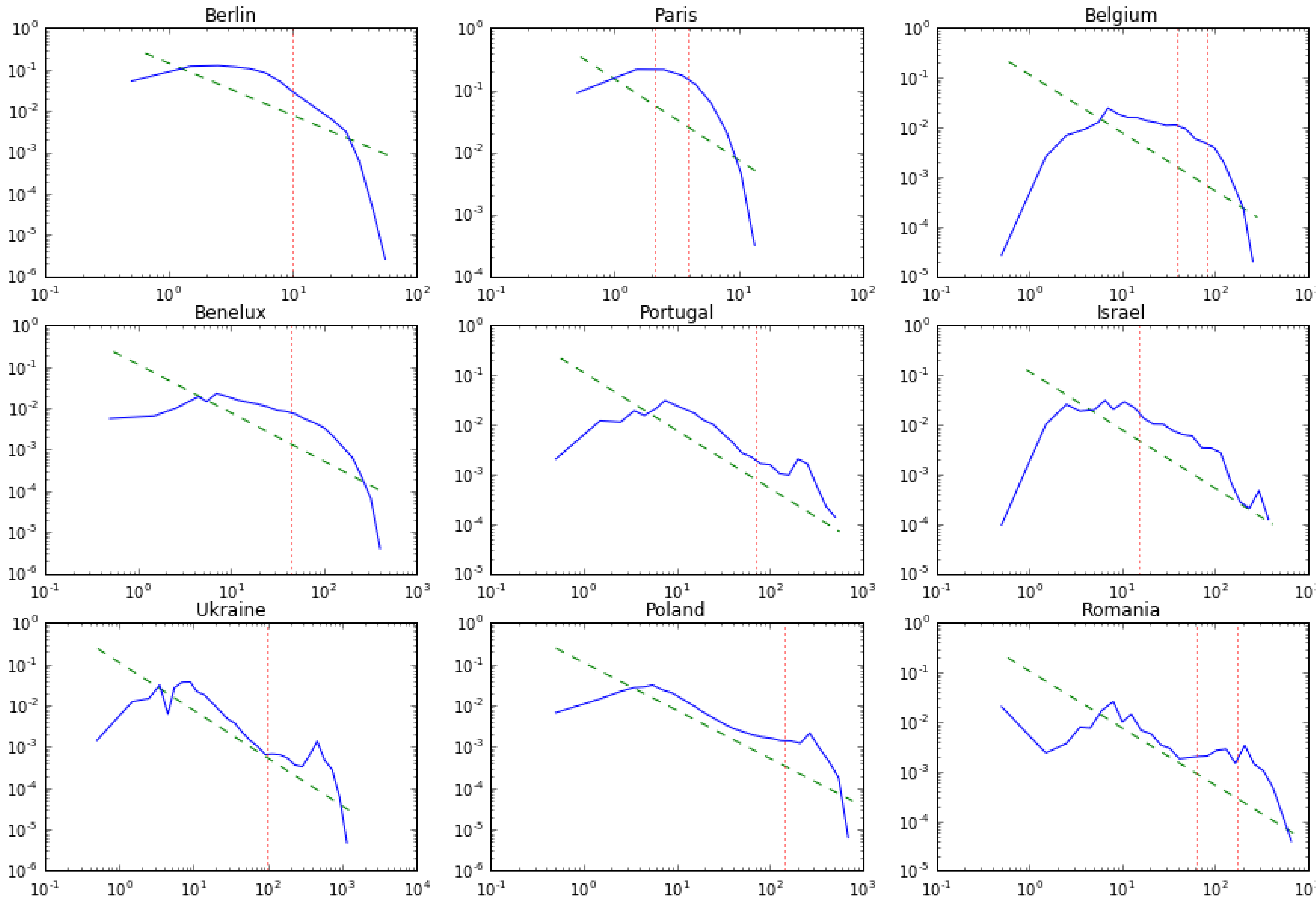
Natural Scales

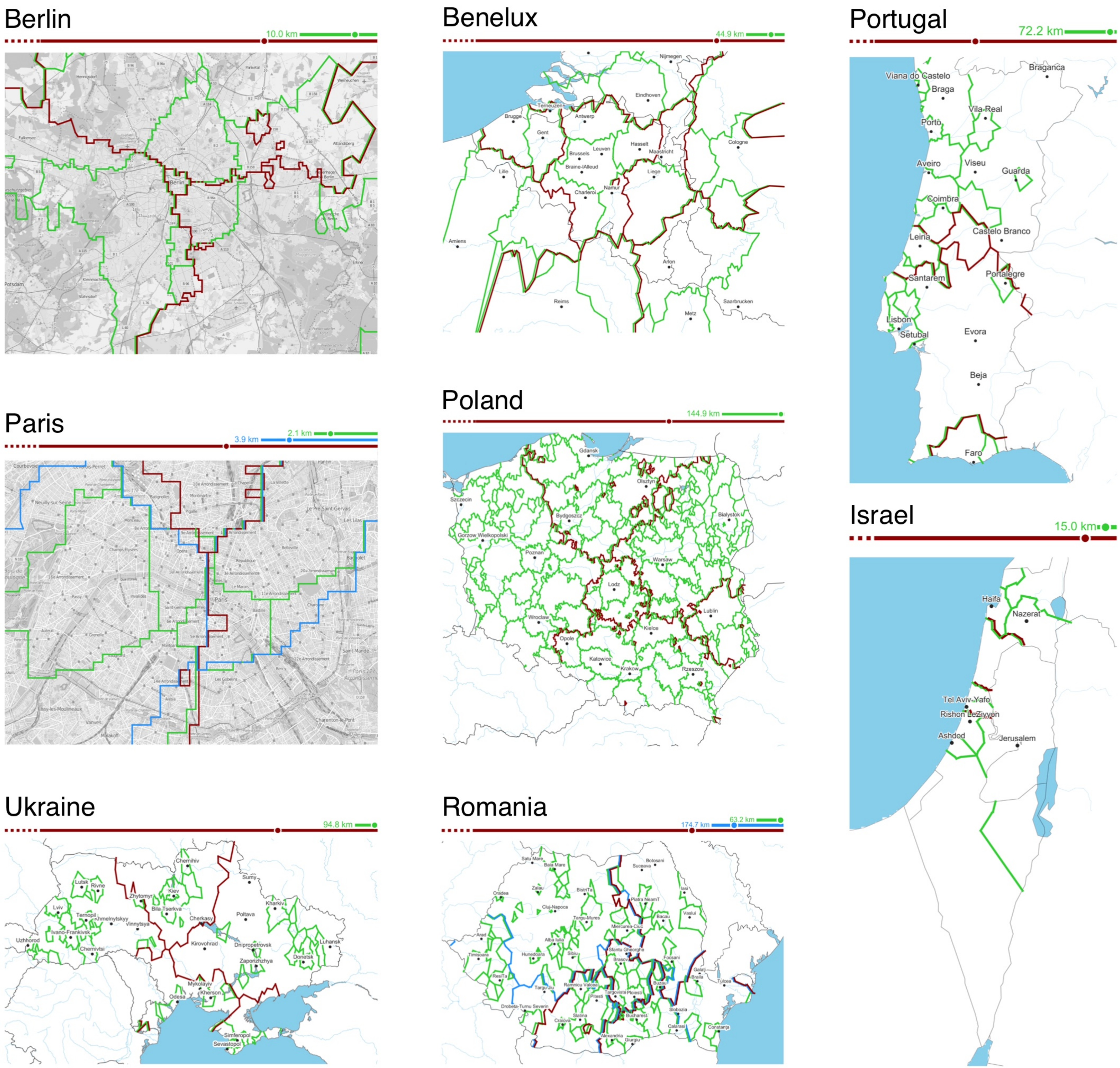
For a given interval I, the natural scale s_I is the percentile of I with the partition that is the most similar to all other partitions in I. To formalize:

$$s_I = \text{argmax}_{s \in I} \sum_{s' \in I} \delta(V, P_s, P_{s'}) \tag{5}$$

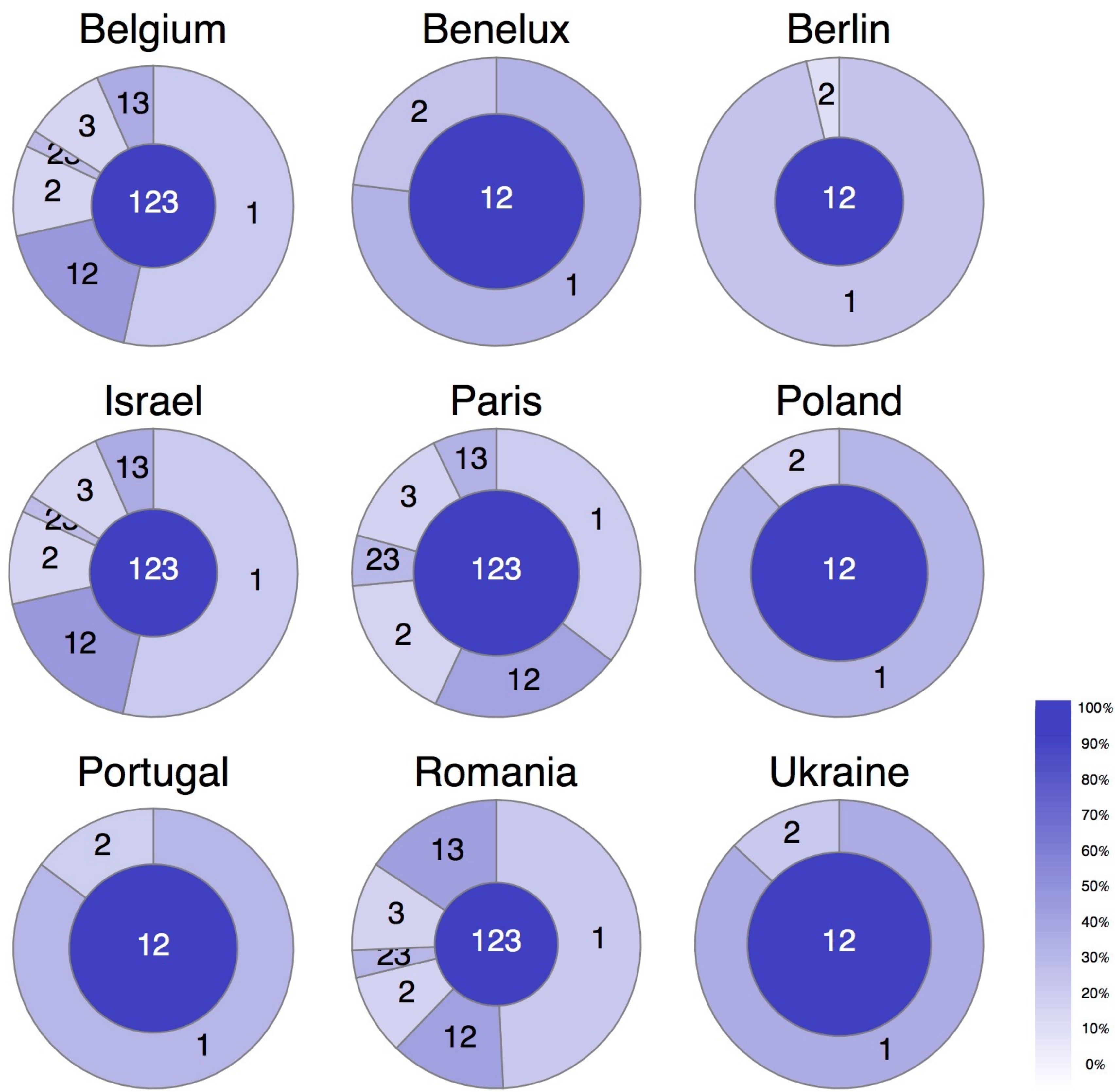

Belgium
Benelux
Berlin
Israel
Paris
Poland
Portugal
Romania
Ukraine
0
20
40
60
80
100

Distance distributions




Berlin
10.0 km
Benelux
44.9 km
Portugal
72.2 km
Viana do Castelo
Braga
Braganca
Vila Real
Porto
Aveiro
Viseu
Guarda
Coimbra
Castelo Branco
Leiria
Santarem
Portalegre
Lisbon
Setubal
Evora
Beja
Faro
Paris
3.9 km
2.1 km
Poland
144.9 km
Israel
15.0 km
Haifa
Nazerat
Tel Aviv-Yafo
Rishon LeZiyyon
Ashdod
Jerusalem
Ukraine
94.8 km
Romania
174.7 km
63.2 km

Scale-related behaviour, not scale-related users



Take Home

- Digital traces on social media can be mined to infer large corpus of human movement data.

- Geographic communities are scale-dependent.

- There are ranges of scales where community partitions remain similar, interspersed with sudden changes (phase transitions).

- It is possible to find the characteristic scales of a region and multi-scale analysis provides insight that would be hidden otherwise.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Natural Scales in Geographical Patterns

Telmo Menezes[1,*] & Camille Roth[1,2,3,*]

Received: 06 October 2016

Accepted: 06 March 2017

Published: xx xx xxxx

Human mobility is known to be distributed across several orders of magnitude of physical distances, which makes it generally difficult to endogenously find or define typical and meaningful scales. Relevant analyses, from movements to geographical partitions, seem to be relative to some ad-hoc scale, or no scale at all. Relying on geotagged data collected from photo-sharing social media, we apply community detection to movement networks constrained by increasing percentiles of the distance distribution. Using a simple parameter-free discontinuity detection algorithm, we discover clear phase transitions in the community partition space. The detection of these phases constitutes the first objective method of characterising endogenous, natural scales of human movement. Our study covers nine regions, ranging from cities to countries of various sizes and a transnational area. For all regions, the number of natural scales is remarkably low (2 or 3). Further, our results hint at scale-related behaviours rather than scale-related users. The partitions of the natural scales allow us to draw discrete multi-scale geographical boundaries, potentially capable of providing key insights in fields such as epidemiology or cultural contagion where the introduction of spatial boundaries is pivotal.