

Soft image segmentation: on the clustering of irregular, weighted, multivariate marked networks

Raphaël Ceré and François Bavaud

Department of Geography and Sustainability
University of Lausanne, Switzerland

Abstract. The contribution exposes and illustrates a general, flexible formalism, together with an associated iterative procedure, aimed at determining soft memberships of marked nodes in a weighted network. Gathering together spatial entities which are both spatially close and similar regarding their features is an issue relevant in image segmentation, spatial clustering, and data analysis in general. Unoriented weighted networks are specified by an “exchange matrix”, determining the probability to select a pair of neighbors. We present a family of membership-dependent free energies, whose local minimization specifies soft clusterings. The free energy additively combines a mutual information, as well as various energy terms, concave or convex in the memberships: within-group inertia, generalized cuts (extending weighted Ncut and modularity), and membership discontinuities (generalizing Dirichlet forms). The framework is closely related to discrete Markov models, random walks, label propagation and spatial autocorrelation (Moran’s I), and can express the Mumford-Shah approach. Four small datasets illustrate the theory.

Keywords : free energy, image segmentation, iterative clustering, soft K-means, Laplacian, modularity, Moran’s I , Mumford-Shah functional, multivariate features, Ncut, soft membership, spatial autocorrelation, spatial clustering

1 Introduction

Regional data analysis, as performed on geographic information systems, deals with a notion of “*where*” (the spatial disposition of regions), a notion of “*what*” (the regional features), and a notion of “*how much*” (the relative importance of regions, as given by their surface or the population size). The data define a *marked, weighted network*, generally *irregular* (think e.g. of administrative units): weighted vertices represent the regions, weighted edges measure the proximity between regions, on which uni- or multivariate features (the marks) are defined.

Much the same can be said of an image made of pixels, that is a collection of elements embedded in a bidimensional layout. The regularity of the setup

(regular grid, uniform weights, binary regular adjacencies) is exploited in most segmentation algorithms, but the latter may become unadapted, precisely, under irregular situations, such as pixels of various sizes or importance, aggregated pixels, irregular boundaries or connectivities, multi-layered or partially missing data.

The fields of spatial analysis, in particular spatial clustering, on one hand, and image segmentation on the other hand, seem currently to be investigated by distinct, non-overlapping scientific communities. Yet, both communities arguably share the same *what-versus-where-trading* challenge, aimed at obtaining clusters both homogeneous and connected.

This contribution proposes a family of iterative algorithms for unsupervised or semi-supervised image segmentation. It attempts to merge a regularized approach to (non-marked) network clustering with the soft K-means of unconnected, feature-marked observations. The underlying weighted formalism, dubbed “ZED formalism” in short (section 2.1), appears to be relevant for spatial analysis, network clustering, and data analysis in general.

In a nutshell, the network clustering objective is expressed by a *generalized cut* functional, encompassing the *Ncut minimisation* (Shi and Malik [36]; Grady and Schwartz [21]) and *modularity maximization* (Newman [30]) as particular cases (section 2.4). This objective is enriched with a features dissimilarity term, central in the *K-means approach*, and further regularized by an *entropy term*. The three terms are additively combined into a freely parametrized *free energy*, whose minimization provides a principled approach, semi-supervised or unsupervised, to the iterative computation of locally optimal solutions, and the emergence of soft clusters (section 3). The soft nature of clusters induces a membership uncertainty, betraying the inter-cluster boundary (section 5.2). Alternatively, pixels can be finally assigned to the group maximizing their membership, thus defining usual hard clusterings.

Replacing the former generalized cut by a *discontinuity functional* constitutes another meaningful option (section 2.3), and both approaches, although opposite regarding their convexity properties (5.1), are investigated and illustrated on four small datasets (section 4).

The quest for good clusterings makes sense for homogeneous enough images, as attested by the standardized value of Moran’s I , the canonical measure of spatial autocorrelation, as defined and illustrated in the present multivariate, weighted setting (section 2.2).

Also, the formal connection with the celebrated Mumford-Shah approach in image segmentation is made explicit (section 5.3), culminating in the identification of a non-additive objective involving both the network structure and the pixel features (equation (19)).

A discussion (section 6) lists some further research lines, and addresses the connection with some alternative approaches to the clustering of marked networks. The appendix details two constructions (diffusive and Metropolis-Hastings) of

the so-called exchange matrix (Berger and Snell [9]), as well as the test of spatial autocorrelation in a weighted, multivariate setting.

The present contribution is an extended version of a paper (Raphaël Ceré [34]) first published in the GISTAM 2017 proceedings. Supplementary material includes the definition and study of *generalized cut functionals*¹, membership uncertainties, Metropolis-Hastings exchange matrices, and new illustrations. It also presents a presumably original *soft Mumford-Shah-like* approach (section 5.3), proposing an adaptation of the original Mumford-Shah image segmentation framework to the “ZED formalism” under investigation (section 2.1).

2 Definitions and formalism

The formalism we consider extends the *spatial autocorrelation formalism* used in Quantitative Geography and Spatial Econometrics to the case of weighted, irregular regions, as well as to multivariate features. It turns out to be extensive enough to provide a flexible framework for unsupervised or semi-supervised *generalized image* segmentation, where the “generalized images” under consideration can be made of irregular pixels, irregularly inter-connected, and endowed with multivariate numerical features.

2.1 The ZED framework

In short, the spatial structure of the network (the “where”) is specified by a square affinity or *exchange matrix* E , interpretable as a joint probability of selecting pairs of pixels; the multi-labelled pixel marks (the “what”) are specified by a square *dissimilarity matrix* D . A soft network clustering is specified by a rectangular *membership matrix* Z , assigning the pixels into groups.

Unsupervised clustering consists in determining a reasonably good Z , taking into account both E (strongly connected pixels should belong to the same group) and D (strongly dissimilar pixels should belong to distinct groups).

In the general, irregular setup, the relative weights $f_i > 0$ of the n elementary vertices, regions or pixels are unequal, but fixed. Their feature dissimilarities D_{ij} are also given, while the specification of E , reflecting the symmetrical network affinity between vertices, enjoys some flexibility (section 7.1), yet reflecting in any case the neighborhood structure of the network, and normalized to $e_{i\bullet} = f_i$ (see below). By contrast, the memberships Z are entirely free, and define a soft clustering of the marked network – whose good enough instances constitute the quest of the present paper.

Space as a weighted network: the exchange matrix E

Specifically, consider n regions (generalized pixels) with relative weights $f_i > 0$, normalized to $f_{\bullet} = \sum_{i=1}^n f_i = 1$, together with an $n \times n$ symmetric non-negative

¹ besides the generalized discontinuity functionals, already addressed in the proceedings, but unfortunately referred there to as “cut functionals”

exchange matrix $E = (e_{ij})$, and *weight-compatible* in the sense $e_{i\bullet} = \sum_{j=1}^n e_{ij} = f_i$. Here and in the sequel, “ \bullet ” denotes the sum over all values of the replaced index.

The exchange matrix E interprets as a joint probability $p(i, j) = e_{ij}$ to select the pair of regions i and j (edges), and defines a weighted unoriented network. Its margins interpret as the probability $p(i) = f_i$ to select region i (vertices).

Weight-compatible exchange matrices E define a continuous neighborhood relation between regions. They can be constructed from f and the adjacency matrix A , or from another spatial proximity of distance matrix (see the appendix). The row-standardized matrix of spatial weights $W = (w_{ij})$ of spatial autoregressive models obtains as $w_{ij} = e_{ij}/f_i$ and constitutes the transition matrix of a reversible Markov chain with stationary distribution f .

Multivariate features: the dissimilarity matrix D

Regional features or marks can consist of univariate grey levels, multivariate color or spectral intensities, or (in a geographical context) any regional variable such as the proportions of specific land uses, population density, proportion of retired people, etc. Multivariate characteristics x_i are suitably combined into $n \times n$ squared Euclidean dissimilarities $D_{ij} = \|x_i - x_j\|^2$.

Soft clustering: the membership matrix Z

A soft *regional clustering* or *image segmentation* into m groups is described by a non-negative $n \times m$ *membership matrix* $Z = (z_{ig})$ with $z_{ig} = p(g|i) \geq 0$ denotes the probability that region (pixel) i belongs to group g , and obeys $z_{i\bullet} = \sum_{g=1}^m z_{ig} = 1$.

The relative weights of the corresponding groups obtain as $\rho_g = \sum_i f_i z_{ig} = p(g) \geq 0$, with $\sum_g \rho_g = 1$. The regional distribution of group g is $f_i^g = p(i|g) = p(g|i)p(i)/p(g) = f_i z_{ig}/\rho_g$, and obeys $f_i^g = 1$.

2.2 Spatial autocorrelation: Moran's I

Obtaining a clustering Z both satisfactory regarding the network E and the features D supposes a kind of compatibility between E and D , and this precisely constitutes the issue of *spatial autocorrelation*, as measured by the weighted, multivariate generalization (2) of Moran's I .

Average multivariate dissimilarities between regions are expressed by *inertias*, generalizing the univariate *variances*. The inertia between randomly selected regions, and the *local inertia* between neighbors, are respectively defined as

$$\Delta = \frac{1}{2} \sum_{i,j=1}^n f_i f_j D_{ij} \quad \Delta_{\text{loc}} = \frac{1}{2} \sum_{i,j=1}^n e_{ij} D_{ij} \quad (1)$$

Comparing the global versus local inertias provides a multivariate generalization of *Moran's I* , namely,

$$I \equiv I(E, D) = \frac{\Delta - \Delta_{\text{loc}}}{\Delta} \quad (2)$$

whose values range in $[-1, 1]$. A large positive I is expected for an image made of large patches characterized with constant features, or at least varying smoothly on average (spatial continuity = positive autocorrelation). A large negative I characterizes an image whose pixel features are contrasted, opposite to their neighbors - such as a chess board with “rook” adjacency. Yet, the value of I in itself is little informative (large values of I are expected whenever diagonal terms E are important), in contrast to its standardized value z , which can furthermore be directly tested by the normal procedure, or by the weight-corrected permutation procedure (section 7.2).

2.3 Image segmentation by generalized discontinuity minimization

The region-group dependency can be measured by the *mutual information*

$$\mathcal{K}[Z] = \sum_{i=1}^n \sum_{g=1}^m p(i, g) \ln \frac{p(i, g)}{p(i)p(g)} = \sum_{ig} f_i z_{ig} \ln \frac{z_{ig}}{\rho_g} \quad (3)$$

A good clustering should consist of homogeneous groups made of regions not too dissimilar regarding their features, that is insuring a low value of the *within-group inertia* (e.g. Bavaud [4])

$$\Delta_W[Z] = \sum_{g=1}^m \rho_g \Delta_g \quad \text{where} \quad \Delta_g = \frac{1}{2} \sum_{ij} f_i^g f_j^g D_{ij} \quad (4)$$

A good clustering should also avoid to separate a pair of spatially strongly connected pixels, that is to insure a low value of the *generalized discontinuity*

$$\mathcal{G}^\kappa[Z] = \sum_{g=1}^m \frac{\varepsilon[z^g]}{\rho_g^\kappa} \quad \text{where} \quad \varepsilon[z^g] = \frac{1}{2} \sum_{ij} e_{ij} (z_{ig} - z_{jg})^2 \quad \text{and} \quad \kappa \in [0, 1] \quad (5)$$

The term $\varepsilon[z^g]$ is called Dirichlet form in potential theory, and attains its minimum value zero iff all pixels lying in a connected component of the network possess the same membership in g .

We consider a *regularized clustering problem*, aiming at determining, among the set \mathcal{Z} of all memberships matrices, a $n \times m$, non-negative and row-normalized matrix Z minimizing the *free energy* functional

$$\mathcal{F}[Z] = \beta \Delta_W[Z] + \frac{\alpha}{2} \mathcal{G}^\kappa[Z] + \mathcal{K}[Z] \quad (6)$$

where $\alpha, \beta \geq 0$. The terms Δ_W , respectively \mathcal{G}^κ , behaves as a features dissimilarity energy, respectively a spatial energy, favoring *hard* partitions obeying $z_{ig} = 0$ or $z_{ig} = 1$. By contrast, the regularizing entropy term \mathcal{K} favors the emergence of soft clusterings. Setting $\alpha = 0$ yields the soft K -means algorithm based on spherical Gaussian mixtures, where the *inverse temperature* β fixes the dissimilarity bandwidth. Canceling the first-order derivative of the free energy with

respect to z_{ig} under the constraints $z_{i\bullet} = 1$ yields the minimization condition

$$z_{ig} = \frac{\rho_g \exp(-H_{ig})}{\sum_h \rho_h \exp(-H_{ih})} \quad (7)$$

where

$$\rho_g[Z] = \sum_{i=1}^n f_i z_{ig} \quad H_{ig}[Z] = \beta D_i^g + \alpha \rho_g^{-\kappa} (\mathcal{L}z^g)_i - \frac{\alpha\kappa}{2} \rho_g^{-\kappa-1} \varepsilon[z^g] . \quad (8)$$

Here $D_i^g = \sum_j f_j^g D_{ij} - \Delta_g$ is the squared Euclidean dissimilarity from i to the centroid of group g , and $(\mathcal{L}z^g)_i = z_{ig} - \sum_j w_{ij} z_{jg} = z_{ig} - (Wz^g)_i$ is the *Laplacian* of membership z^g at pixel i , comparing its value to the average value of its neighbors, and adjusting the former to the latter. For $\kappa > 0$, this adjustment mechanism is downscaled for large groups (factor $\rho_g^{-\kappa}$); in addition, spatially discontinuous small groups are encouraged to grow by the last term in (8), independent of i .

2.4 Image segmentation by generalized cut minimization

Another functional whose minimization favors spatially connected clusters is the *generalized cut*

$$\mathcal{C}^\kappa[Z] = \sum_{g=1}^m \frac{\rho_g^2 - e(g,g)}{\rho_g^\kappa} \quad \text{where} \quad e(g,g) = \sum_{i,j=1}^n e_{ij} z_{ig} z_{jg} \quad \text{and} \quad \kappa \in [0, 1] . \quad (9)$$

The choice $\kappa = 1$ amounts to the *N-cut objective* (Shi and Malik [36]), while the choice $\kappa = 0$ is equivalent to the *modularity criterium* (Newman [30]). Again, minimizing the corresponding free energy

$$\mathcal{F}[Z] = \beta \Delta_W[Z] + \frac{\gamma}{2} \mathcal{C}^\kappa[Z] + \mathcal{K}[Z] \quad \beta, \gamma \geq 0 \quad (10)$$

yields the necessary first-order condition (7), where

$$H_{ig}[Z] = \beta D_i^g + \gamma \rho_g^{-\kappa} [\rho_g - (Wz^g)_i] - \frac{\gamma\kappa}{2} \rho_g^{-\kappa-1} [\rho_g^2 - e(g,g)] . \quad (11)$$

For $\kappa = 0$ (modularity clustering), the term $\rho_g - (Wz^g)_i$ compares the average membership of the neighbors of i to the overall average membership (rather than to the membership of i itself, as in (8)). The term $(Wz^g)_i$ precisely implements the *label propagation* mechanism acting in some network clustering algorithms (e.g. Zhu and Ghahramani [41], Raghavan et al. [33]).

For $\kappa > 0$, this adjustment mechanism is downscaled for large groups (factor $\rho_g^{-\kappa}$); in addition, loosely intra-connected small groups are encouraged to grow by the last term in (11), independent of i , which prevents the creation of clusters made out of a single pixel – a known defect of the unnormalized cut criterium (e.g. von Luxburg [28]).

3 Iterative procedure: unsupervised and semi-supervised

Equation (7) can be solved iteratively from some initial membership $Z^0 \in \mathcal{Z}$, updating at each step ρ_g and the exponent H_{ig} in versions (8) or (11), until convergence to Z^∞ , which constitutes a *local minimum* of $\mathcal{F}[Z]$. Matrix Z^∞ constitutes the searched for soft spatial partition or image segmentation of the unsupervised procedure. It can be further hardened by assigning each pixel i to group $g = \arg \max_h z_{ih}^\infty$, breaking possible ties at random.

A semi-supervised implementation of the procedure, imposing the membership of a few pixels (and possibly breaking down the monotonic decrease of $\mathcal{F}[Z]$: see figures 3 to 8) goes as follow: first, the set Ω of the n regions is partitioned into two disjoint, non-empty sets, namely the user-defined *tagged regions* T , and the *free regions* F , with $\Omega = T \cup F$ and $T \cap F = \emptyset$. The tagged set T itself consists of m non-empty disjoint subregions $T = \cup_{\tau=1}^m T_\tau$ initially tagged with m distinct *strokes* applied on a small number of pixels: they form the *seeds* of the $g = 1, \dots, m$ figures to be extracted, while the remaining regions will be assigned to the background numbered $g = 0$.

Memberships $Z = (z_{ig})$ consist of $n \times (m+1)$ non-negative matrices obeying $\sum_{g=0}^m z_{ig} = 1$. Their initial value Z^0 is set as

$$z_{ig}^0 \begin{cases} = 1 & \text{if } i \in F \text{ and } g = 0 \\ = 1 & \text{if } i \in T_\tau \text{ and } g = \tau \\ = 0 & \text{otherwise .} \end{cases} \quad (12)$$

Iteration (7) is then performed. At the end of each loop, the tagged regions are reset to their initial values $z_{i\tau}^0 = 1$ for all $i \in T_\tau$. After convergence, one expects the hardened clusters obtained by assigning i to group $g = \arg \max_{h=0}^m z_{ih}^\infty$ to consist of m connected figures $g = 1, \dots, m$ each containing the tagged set T_g , as well as a remaining background supported on F .

The iterative image segmentation algorithm summarized below (Table 1) requires

- 1) a vector of n weights $f_i > 0$ associated to each pixel or region
- 2) a vector of n grey levels or multivariate characteristics x_i
- 3) a $n \times n$ binary, symmetric, off-diagonal adjacency matrix A
- 4) a set of disjoint non-empty tagged sets of pixels $\{T_\tau\}_{\tau=1}^m$.

Begin

Fix $\beta > 0$ (the parameter conjugate to the within-clusters inertia) and $\kappa \in [0, 1]$

Compute the weight vector f ($f_i = 1/n$ for regular grids)

Compute the binary adjacency matrix A

diffusive exchange matrix : For a given $t > 0$, compute $E(f, A, t)$ by (20)

Metropolis-Hastings exchange matrix: Compute E by (21)

Compute the matrix of spatial weights as $w_{ij} = e_{ij}/f_i$

Compute the features dissimilarity matrix $D_{ij} = \|x_i - x_j\|^2$

Initialize the $n \times (m + 1)$ membership matrix Z^0 as:

$$z_{ig}^0 = 1 \text{ if } i \in F \text{ and } g = 0$$

$$z_{ig}^0 = 1 \text{ if } i \in T_\tau \text{ and } g = \tau$$

$$z_{ig}^0 = 0 \text{ otherwise.}$$

Loop : $Z^{(r+1)}$ for the r -th iteration, stop after convergence

$$\text{Group weight : } \rho_g = \sum_i f_i z_{ig}$$

$$\text{Emission probabilities } f_i^g = \frac{f_i z_{ig}}{\rho_g}$$

$$\text{Dissimilarity to the centroid : } \hat{D}_i^g = \sum_j f_j^g D_{ij} - \Delta_g$$

generalized discontinuity segmentation : Fix the parameter $\alpha > 0$ and compute H_{ig} by (8)

generalized cut segmentation : Fix the parameter $\gamma > 0$ and compute H_{ig} by (11)

Compute $z_{ig}^{(r+1)}$ by (7)

for the semi-supervised case : Re-initialize $z_{ig}^{(r+1)} = \delta_{g\tau}$ for $i \in T_\tau$

for the unsupervised case : Do nothing

Attribute $i \in F$ to the group $g = \arg \max_{h=0}^m z_{ih}^{(\infty)}$

End

Table 1. Variants of the semi-supervised and the unsupervised iterative segmentation algorithm.

4 Illustrations

4.1 Swiss federal votes

The irregular network consists on the $n = 309$ communes of canton of Vaud, endowed with their diffusive exchange matrix (20) with $t = 1$, where A is the “queen” adjacency matrix and the non-uniform weights f are the proportion of inhabitants. Features consist, for each commune i , of the proportion of “yes” for three Swiss federal initiatives submitted to the citizens on February the 12th 2017, namely (figure 1) x_i for the *Corporate Tax Reform Act III* (refused by 40.9% of voting citizens), y_i for the *Federal Decree on the Simplified Naturalisation of Third-Generation Immigrants* (accepted at 60.4%) and z_i for the *Federal Decree on Establishing a Fund for National Roads and Urban Traffic* (accepted at 61.9%). Dissimilarities are simply defined as $D_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2$, further rescaled in the range $[0, 1]$.

Figures 3 to 8 depict the semi-supervised hard assignment obtained from the initial strokes $T_1 = \{20\}$ (Bière; group 1), and $T_2 = \{3\}$ (Chessel; group 2), after 100 iterations, together with the change of the free energy $\mathcal{F}[Z]$ during the iteration. Generalized cuts and discontinuity variants are both tested for various values of the parameters κ , β , γ and α . In particular, the conditional entropy $H(G|O)$, measuring the softness of the partition (section 5.2), *decreases in β* , *decreases in γ* , but *increases in α* , as expected.

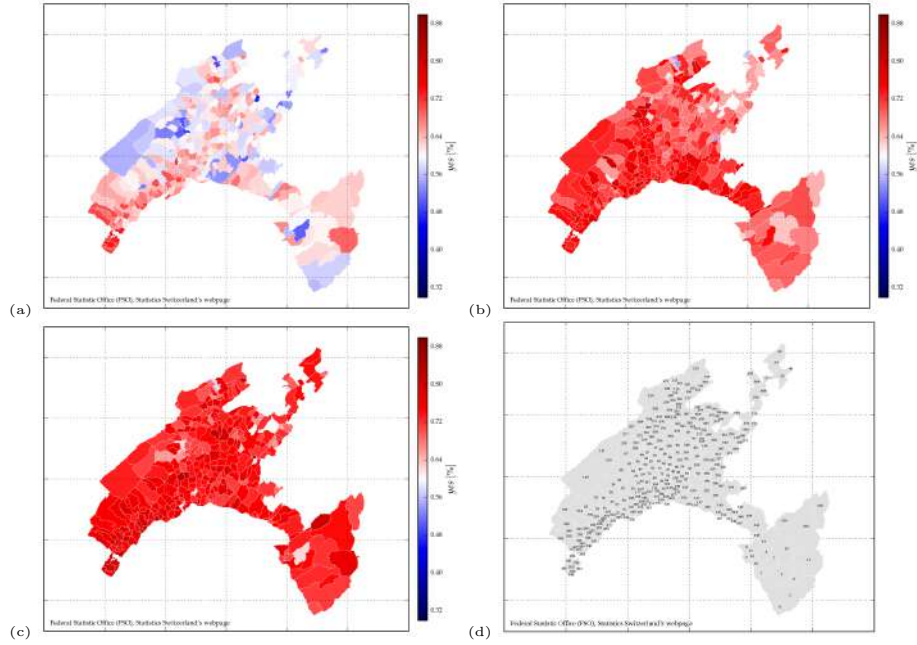


Fig. 1. *Swiss federal votes, canton of Vaud:* proportion of “yes” (a) x_i for the first votation, (b) y_i for the second votation, and (c) z_i for the third votation. (d) lists the commune numbers used by the cantonal administration.

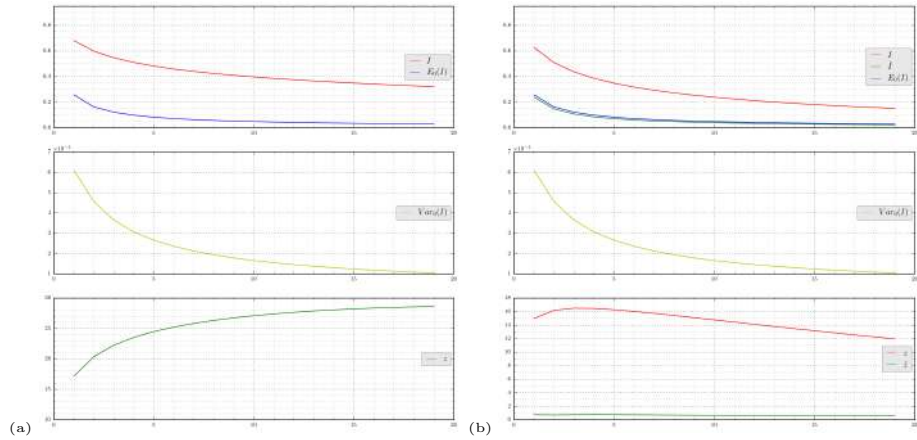


Fig. 2. *Swiss federal votes, spatial autocorrelation:* (a) Moran’s I (top) and standardized z normal test value (bottom), as a function of the free parameter $t \in [1, 20]$ of the diffusive exchange matrix specification (20). (b) Moran’s I after a plain permutation of the political features (x, y, z) between communes, and Moran’s \hat{I} after applying the weight-corrected permutation (22). Note the latter to almost coincide with the expected value $E_0(I)$ under the null hypothesis, as it must.

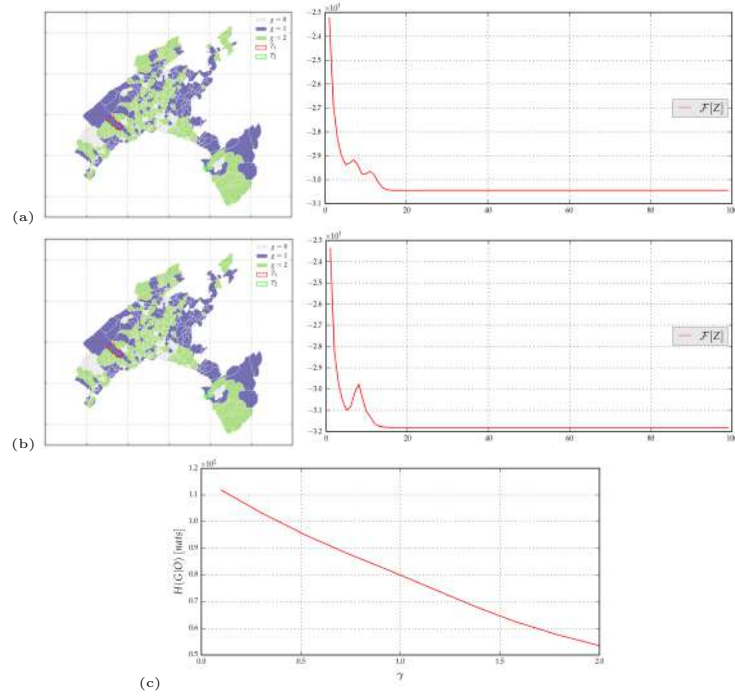


Fig. 3. Swiss federal votes, continued: **cut minimization** with $\kappa = 0.0$, $\beta = 300$, (a) $\gamma = 0.1$ and (b) $\gamma = 2.0$. (c) decrease of the conditional entropy in γ .

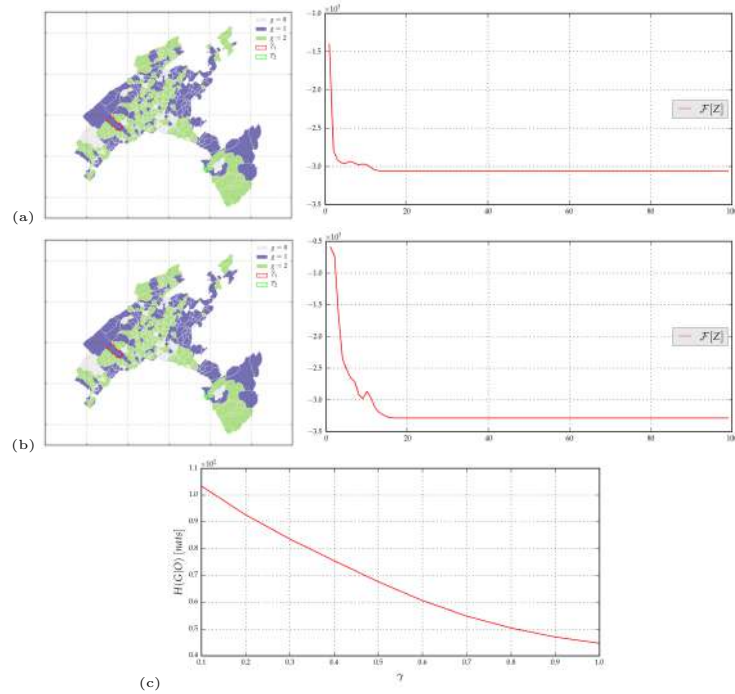


Fig. 4. Swiss federal votes, continued: **cut minimization** with $\kappa = 1.0$, $\beta = 300$, (a) $\gamma = 0.1$ and (b) $\gamma = 1.0$. (c) decrease of the conditional entropy in γ .

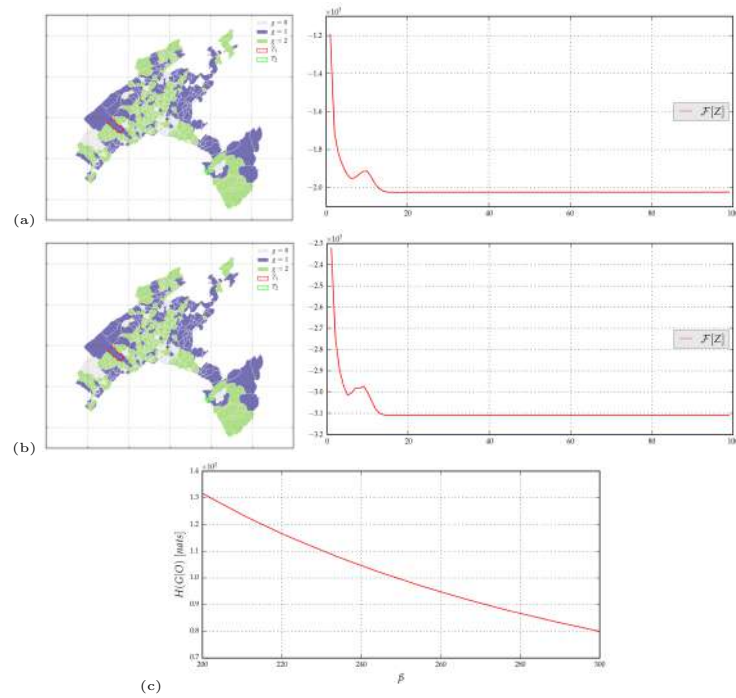


Fig. 5. Swiss federal votes, continued: **cut minimization** with $\kappa = 0.0$, $\gamma = 1.0$, (a) $\beta = 200.0$ and (b) $\beta = 300.0$. (c) decrease of the conditional entropy in β .

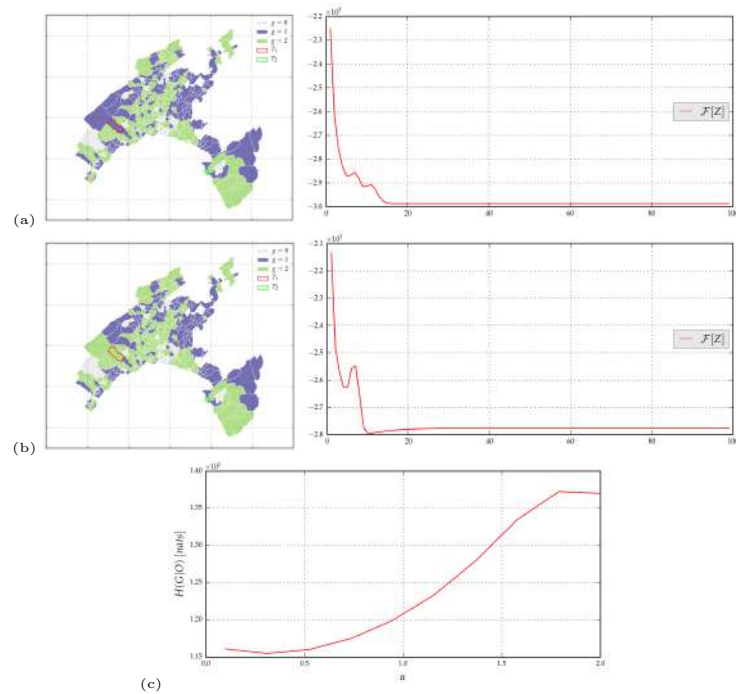


Fig. 6. Swiss federal votes, continued: **discontinuity minimization** with $\kappa = 0.0$, $\beta = 300$, (a) $\alpha = 0.1$ and (b) $\alpha = 2.0$. (c) increase of the conditional entropy in α .

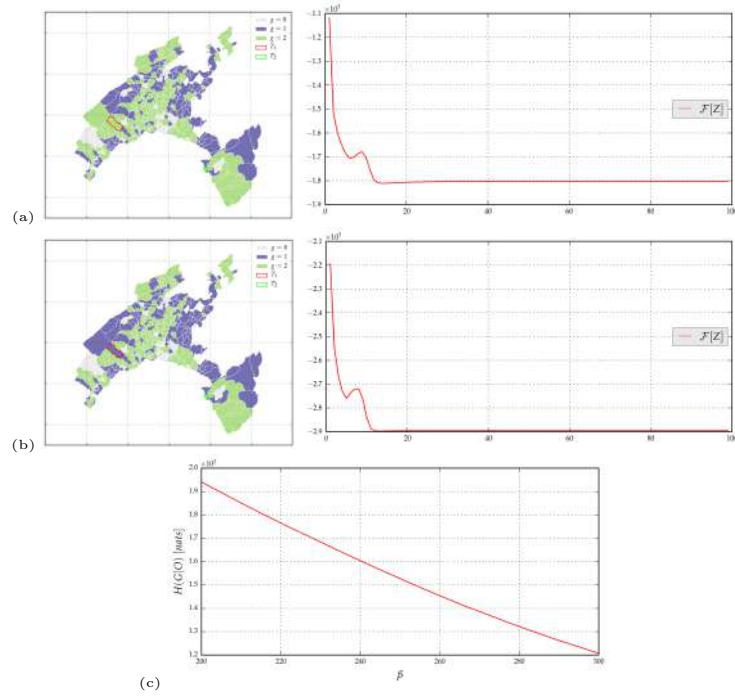


Fig. 7. *Swiss federal votes, continued: discontinuity minimization* with $\kappa = 0.0$, $\alpha = 1.0$, (a) $\beta = 200$ and (b) $\beta = 300$. (c) decrease of the conditional entropy in β .

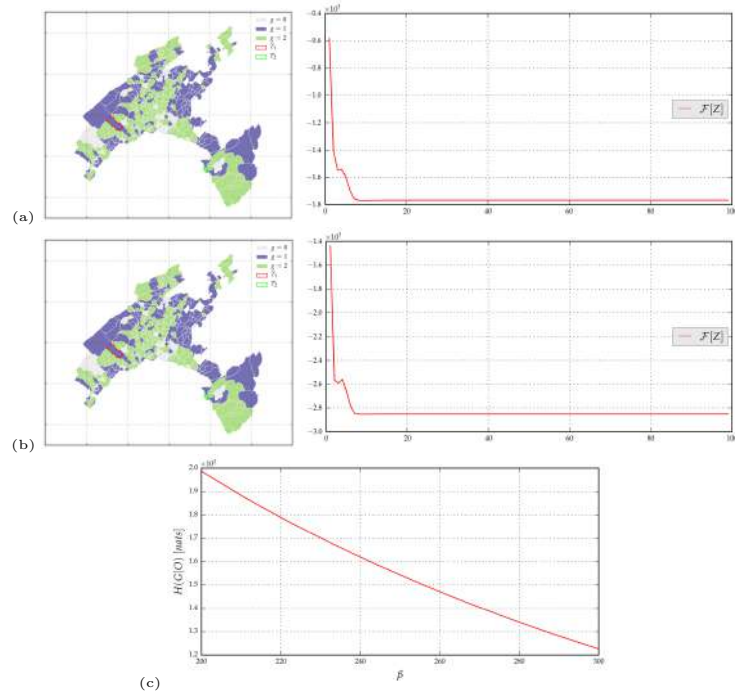


Fig. 8. *Swiss federal votes, continued: discontinuity minimization* with $\kappa = 1.0$, $\alpha = 0.5$, (a) $\beta = 200$ and (b) $\beta = 300$. (c) decrease of the conditional entropy in β .

4.2 The Portrait

Figure 9 refers to a regular trivariate image (levels of red green blue) of size $n = 100 \times 115$ with uniform weight vector $f_i = 1/n$. Again, the binary adjacency matrix A has been built under the “queen” scheme (8 neighbors for inside pixels), on which the Metropolis-Hastings exchange matrix (21) have been adopted. D_{ij} is the sum of the squared differences between color intensities, rescaled to $[0, 1]$. The corresponding Moran’s I (2) is 0.52 for an expectation of $E_0(I) = 0.008$ and variance $\text{Var}_0(I) = 2.201 \times 10^{-05}$ (section 7.2). Normal test value $z = 108.48$ denotes a massively significant spatial autocorrelation, as it must.

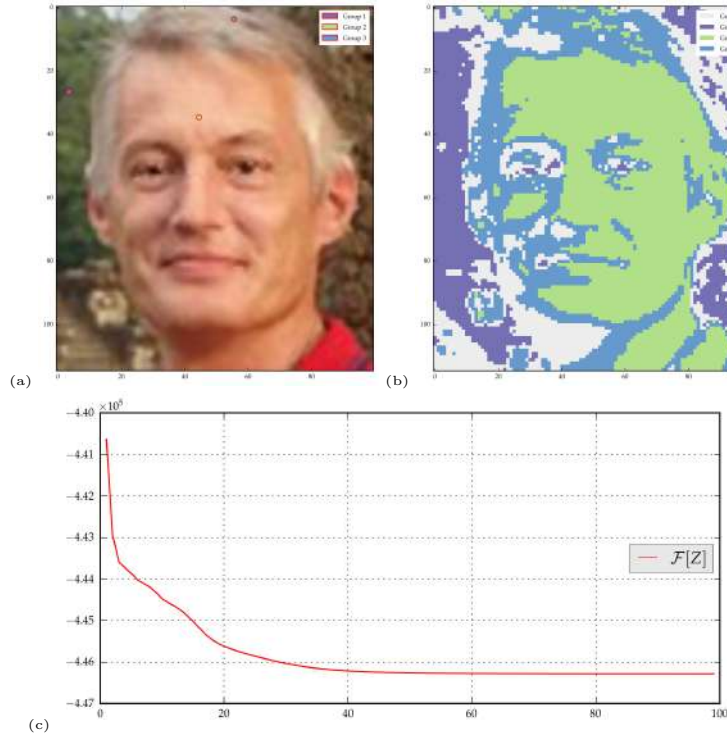


Fig. 9. *The Portrait:* (a-b) original image with initial strokes and unsupervised hard assignment obtained for background group 0 colored in grey, group 1 in purple, group 2 in green and group 3 in blue, after 100 iterations; (c) decrease of the free energy $\mathcal{F}[Z]$ during the iteration for a **discontinuity minimization** with $\kappa = 0.0$, $\alpha = 0.1$ and $\beta = 950.0$.

4.3 The Geometer

Figure 10 refers to the levels of grey of a regular image of size $n = 50 \times 50$. The node weights f and univariate dissimilarities D are constructed as above but the diffusive exchange matrix (20) has been adopted. The corresponding Moran’s I

is 0.848, with an expectation of $E_0(I) = 0.0174$, variance $\text{Var}_0(I) = 0.0001$ and test value $z = 81.88$.

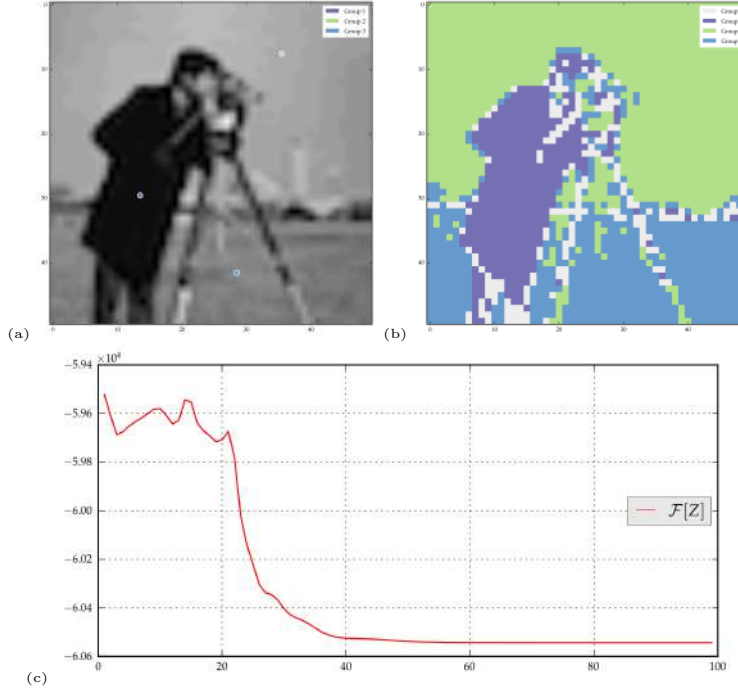


Fig. 10. The Geometer: (a-b) original image with initial strokes and unsupervised hard assignment obtained for background group 0 colored in grey, group 1 in purple, group 2 in green and in blue, after 100 iterations; (c) decrease of the free energy $\mathcal{F}[Z]$ during the iteration for a **generalized cut minimization** with $\kappa = 0.0$, $\gamma = 4.0$ and $\beta = 350.0$.

4.4 Lausanne

One considers the rectangular network made of $n = 50 \times 60$ hectometers (census blocks) in the region of Lausanne, Switzerland, with regular queen binary adjacency matrix A . The features count, for each hectometer, the number of built units possessing a given characteristic, among $p = 63$ criteria in 2014 (source: Swiss Federal Statistical Office). The corresponding squared Euclidean dissimilarities are rescaled to $[0, 1]$.

Figure 11 refers to a uniform weight $f_i = 1/n$, and figure 12 to a non-uniform weight f proportional to the number of inhabitants in each hectometer. Note the diffusive scheme (21) to generate two distinct exchange matrices: A is the same, but f differs. One finds $I = 0.404$, $E_0(I) = 0.258$, $\text{Var}_0(I) = 3.019 \times 10^{-5}$ and $z = 26.550$ for the uniform case; by contrast, $I = 0.877$, $E_0(I) = 0.211$, $\text{Var}_0(I) = 0.0002$ and $z = 52.943$ for the non-uniform case.

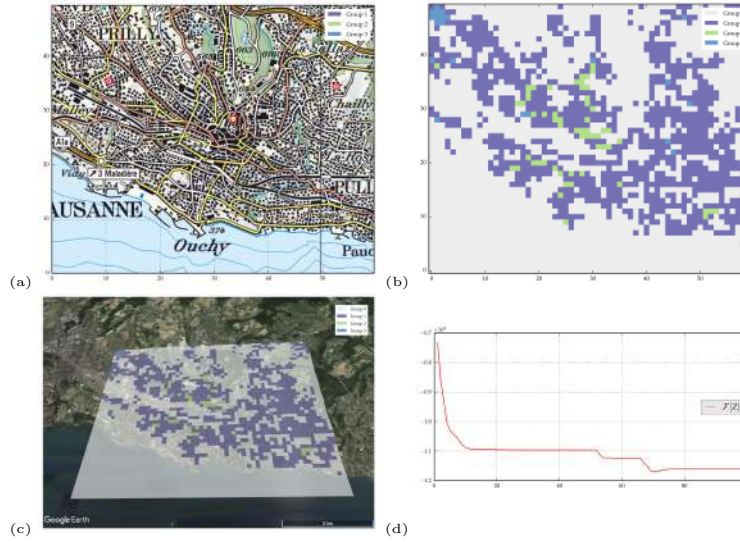


Fig. 11. Lausanne, uniform case: (a-b) topographic map with initial strokes of the area under study and unsupervised hard assignment obtained for background group 0 colored in grey, group 1 in purple, group 2 in green and group 3 in blue, after 100 iterations; (c) hard assignment projected (d) decrease of the free energy $\mathcal{F}[Z]$ during the iteration for a **generalized cut minimization** with $\kappa = 0.0$, $\gamma = 1.0$ and $\beta = 250.0$.

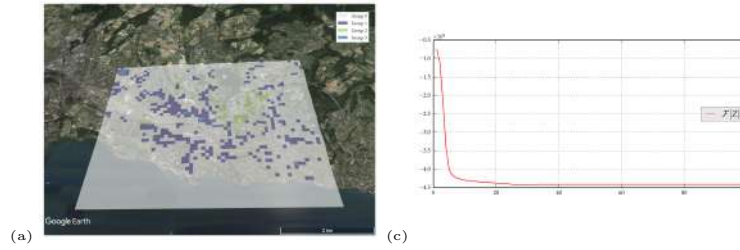


Fig. 12. Lausanne, non-uniform case: (a) unsupervised hard assignment projected obtained for background group 0 colored in grey, group 1 in purple, group 2 in green and group 3 in blue, after 100 iterations; (c) decrease of the free energy $\mathcal{F}[Z]$ during the iteration for a **generalized cut minimization** with $\kappa = 0.0$, $\gamma = 1.0$ and $\beta = 650.0$.

5 Further formal considerations

5.1 Convexity, concavity, and local minima

The set \mathcal{Z} of all memberships is convex, and the within-group inertia $\Delta_W[Z]$ is *concave* in Z (e.g. Bavaud [4]). As a result, the minimum $\arg \min_{Z \in \mathcal{Z}} \Delta_W[Z]$ is attained on the extreme points of \mathcal{Z} , that is on *hard memberships*, whose determination is notoriously difficult. By contrast, the mutual information $\mathcal{K}[Z]$ is *convex*, and attains its minimum zero on “independent” soft memberships of

the form $z_{ig} = \rho_g$. Mixing the two functionals as in the Gaussian mixture model $\beta\Delta_W[Z] + \mathcal{K}[Z]$ generates, for β large enough, a functional possessing many *local minima*, attained after convergence of the iterative procedure of table 1 (in the unsupervised case, and in absence of spatial terms).

The generalized discontinuity functional $\mathcal{G}^\kappa[Z]$ (5) can be shown to be *convex* in Z , for any $\kappa \in [0, 1]$ (proofs are postponed in a forthcoming, more technical note). For a connected network (i.e. whose E exchange matrix is irreducible) the minimizers are again the independent soft memberships $z_{ig} = \rho_g$.

By contrast, the cut functional $\mathcal{C}^\kappa[Z]$ (9) can be shown to be *concave* for any $\kappa \in [0, 1]$, at least for the *positive semi-definite* (p.s.d.) exchange matrices E , that is whose eigenvalues are non-negative; such networks are referred to as *diffusive*.

Diffusive exchange matrices (20) define a diffusive network, precisely, by contrast to Metropolis-Hastings exchange matrices (21) which are not p.s.d. in general. Diffusive networks necessarily possess diagonal components (loops), unlike normalized adjacency matrices, or bipartite graphs. Also, exchange matrices of the “radial basis” form $e_{ij} = a_i a_j \exp(-\lambda d_{ij}^2)$, where d_{ij} is a spatial Euclidean distance and $a_i \geq 0$ a calibrating factor ensuring $e_{i\bullet} = f_i$, are diffusive (see e.g. Bavaud [6]).

For a diffusive network, the membership minimizing $\mathcal{C}^\kappa[Z]$ constitutes a hard partition, difficult to compute, for which various heuristics (such as spectral clustering or label propagation) have been devised. The present regularized approach constitutes another, in line with model-based clustering or simulated annealing (e.g. Rose et al. [35]).

5.2 Clustering softness

The regularizing effect of convex functionals, namely the mutual information $\mathcal{K}[Z]$ and the generalized discontinuity $\mathcal{G}^\kappa[Z]$, are responsible for the softness of the optimal membership Z . Denoting by G , resp. O , the variables “group”, resp. “pixel”, the pointwise conditional entropy $H(G|i) = -\sum_g z_{ig} \ln z_{ig}$ measures the membership uncertainty of pixel i . The quantity $H(G|i)$ is large for pixels located at the group frontiers, and thus provides a possibly original *boundary detection mechanism* (figure 13). Its average $H(G|O) = \sum_i f_i H(G|i) = H(G, O) - H(O)$ constitutes a measure of *overall softness* of the clustering, related to the mutual information as

$$\mathcal{K}[Z] = H(G) + H(O) - H(G, O) = H(G) - H(G|O) \quad (13)$$

In soft K-means, increasing β increases the influence of the (concave) energy Δ_W relatively to the (convex) entropy, and hence decreases the softness $H(G|O)$ of the clustering. By the same reasoning, and in view of the remarks of section 5.1, one expects the softness $H(G|O)$ to *decrease* in γ (cut minimization), but to *increase* in α (discontinuity minimization), as observed in figures 3 to 8.

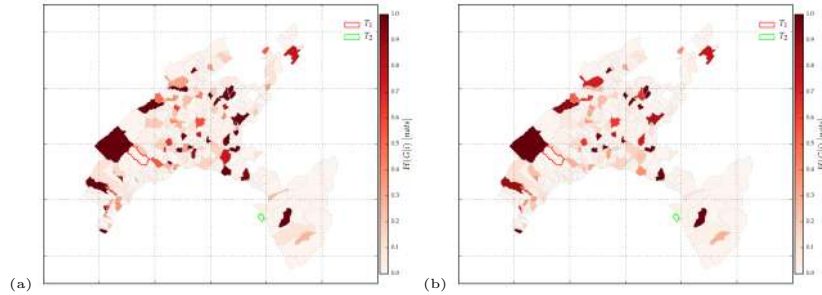


Fig. 13. Conditional pointwise entropy $H(G|i)$ for the conditions of figure 3 with $\kappa = 0.0$, $\beta = 300$, (a) $\gamma = 0.1$ and (b) $\gamma = 2.0$. A large value of $H(G|i)$ denotes a large uncertainty in the membership of region i , lying at some boundary between groups.

5.3 A soft Mumford-Shah-like approach

The influential *Mumford-Shah* approach (Mumford and Shah [29]) aims at governing the image segmentation or “morphogenesis” (e.g. Petitot [32]) of an image whose support $\Omega \subset \mathbb{R}^2$ and “image intensity” $x(s)$ at $s \in \Omega$ are given.

The approach consists in dividing Ω into m disconnected classes Ω_g separated by an inter-classes boundary $B \subset \Omega$ (that is $\Omega \setminus B = \cup_{g=1}^m \Omega_g$) of size $\mathcal{B} = |B|$, as well as by approximating $x(s)$ within each class by a smooth function $u(s)$. The Mumford-Shah functional, to be minimized over u and B , expresses in its original continuous setup as

$$\mathcal{M}(u, B) = \nu \int_{\Omega} (x(s) - u(s))^2 ds + \frac{\delta}{2} \sum_{g=1}^m \int_{\Omega_g} \|\nabla u(s)\|^2 ds + \lambda \mathcal{B} \quad (14)$$

and its rigorous mathematical treatment is fairly demanding (see e.g. Vitti [38]).

In the so-called *cartoon limit* $\delta \rightarrow \infty$, minimization of (14) requires $u(s)$ to be a constant u_g within each hard class Ω_g , with solution $u_g = \bar{x}_g = \int_{\Omega_g} x(s) ds / |\Omega_g|$: color levels are constant within each cell, which are separated by lines of same thickness, and smooth enough to ensure a low value of \mathcal{B} . This definition appears to closely characterize the style of drawing used in Franco-Belgian comics known as *ligne claire* (figure 14), epitomized by the series “The Adventures of Tintin” by Hergé (Gaumer and Moliterni [19]).

In absence of clustering, that is with a single group ($m = 1$, $B = \emptyset$ and $\mathcal{B} = 0$), the discrete approximation of (14) reads

$$\mathcal{M}(u) = \nu \sum_i f_i (x_i - u_i)^2 + \frac{\delta}{2} \sum_{ij} e_{ij} (u_i - u_j)^2 \quad (15)$$

with unique minimizer $u = \mu[(1 + \mu)I - W]^{-1}x$, where $\mu = \nu/\delta > 0$.

Soft partitions Z can be introduced by requiring the signal approximation to be of the form $u_i = \sum_g z_{ig} y_g$, that is $u = Zy$ for some freely adjustable vector



Fig. 14. An example of the *ligne claire* style of drawing: comic mural (2000), Rue des Alexiens 49, Brussels: *Le jeune Albert* by Yves Chaland.

y with m components. They define a *soft, discrete Mumford-Shah functional* of the form

$$\mathcal{M}(Zy, B) = \nu \mathcal{V}[y, Z] + \lambda \mathcal{B}[Z]$$

where

$$\begin{aligned} \mathcal{V}[y, Z] &= \sum_i f_i(x_i - (Zy)_i)^2 + \frac{1}{2\mu} \sum_{ij} e_{ij}((Zy)_i - (Zy)_j)^2 \\ &= (Zy - x)' \Pi (Zy - x) + \frac{1}{\mu} (Zy)' (\Pi - E) Zy \end{aligned} \quad (16)$$

and $\Pi = \text{diag}(f)$ is the diagonal matrix containing the pixel weights. Minimizing (16) over y is an exercise in matrix calculus, and yields the solution

$$y_0 = \Gamma^{-1} \xi \quad \Gamma = Z' T Z \quad T = \frac{1}{\mu} [(1 + \mu) \Pi - E] \quad \xi = Z' \Pi x \quad (17)$$

Plugging y_0 into (16) yields

$$\begin{aligned} \mathcal{V}[Z] &= \min_y \mathcal{V}[y, Z] = \mathcal{V}[y_0, Z] = x' (\Pi - \Pi Z \Gamma^{-1} Z' \Pi) x = \\ &= \frac{1}{2} \text{trace}(\Pi Z (Z' T Z)^{-1} Z' \Pi D) \end{aligned} \quad (18)$$

where $D = (D_{ij})$ denotes the matrix of squared Euclidean dissimilarities between pixel intensities $D_{ij} = (x_i - x_j)^2$, and identity $\Gamma \mathbf{1} = \rho$ has been used in the last expression.

In summary, the original continuous Mumford-Shah functional (14) appears to be *expressible into the present discrete, weighted setting, involving soft partitions Z of marked networks*. Optimal clusterings minimize the functional

$$\mathcal{M}[Z] = \nu \mathcal{V}[Z] + \lambda \mathcal{B}[Z] \quad \nu, \lambda > 0 \quad (19)$$

where $\mathcal{V}[Z]$, given by (18), both depends on the network structure E (through T in (17)) and on the node dissimilarities D , but in a non-additive way, in contrast to proposals (6) and (10). Also, $\mathcal{V}[Z]$ is homogeneous of degree zero (recall that a functional $\mathcal{A}[Z]$ is homogeneous of degree k if, formally, $\mathcal{A}[cZ] = c^k \mathcal{A}[Z]$), while $\mathcal{G}^\kappa[Z]$ and $\mathcal{C}^\kappa[Z]$ are homogeneous of degree $2 - \kappa$, and $\mathcal{K}[Z]$ and $\Delta_W[Z]$ are homogeneous of degree 1.

The inter-cluster boundary length $\mathcal{B}[Z]$ remains to be specified, possibly such as the discontinuity $\mathcal{G}^\kappa[Z]$, or as the generalized cut $\mathcal{C}^\kappa[Z]$, or some other functional. In addition, an entropic regularizing term $\mathcal{K}[Z]$ can be added to (19), hence yielding an alternative, original iterative scheme to be compared to the procedure of section 3.

Expression (18) turns out to hold for multivariate features as well. The issue of the convex versus concave nature of $\mathcal{V}[Z]$, as well as further formal and empirical investigations on soft segmentation based on (19) are postponed for a further work.

In summary, the “ZED formalism” sketched in section 2.1 can express the Mumford-Shah approach within the present framework, whose locally optimal clusters are however bound to differ from the relaxed generalized discontinuity and generalized cut approaches of sections (2.3) and (2.4).

6 Discussion

The formalism presented here expresses and illustrates a few alternatives defining clustering in two close but little interacting domains, namely image segmentation and regional partitioning. Locally optimal soft clusters Z are both adapted to a given irregular unoriented network E on one hand, and to a given set of node dissimilarities D , reflecting the multivariate node features X (marks), on the other hand.

Its basic ingredients have been developed for decades in large, robust scientific communities. In view of the sheer size of the image segmentation domain (including reviews and surveys thereof; see in particular Couprie et al. [13] and Fouss et al. [17]), specific claims of originality seem foolish.

Yet, beside the choice of functionals and parameters, let us underline the flexibility of the approach: the determination of a weight-compatible exchange matrix E reflecting the spatial proximity is a vast issue in itself, covering in large part the theory of discrete reversible Markov chains. Considering *probabilities on paths* (Bavaud and Guex [8]; Françoisse et al. [18]) instead of probabilities on nodes or pairs of nodes, permits to extend the formalism to random walks based modularities (Devooght et al. [15]) or multi-target based clustering (e.g. Sinop and Grady [37]; Guex [24]). This line of research pursues the “electric interpretation” of reversible Markov chains (Doyle and Snell [16]), involving Dirichlet differential equations and computation of the electric potentials, already standard in image segmentation (Grady [20]). The possibility, in probabilistic formulations of random walks, to set *independently* the edge *capacities* and the edge

resistances (Bavaud and Guex [8]; Guex [24]; Fouss et al. [17]), seems especially relevant for the clustering of *marked networks*: it is tempting to identify the capacity contribution as a spatial term enabling transitions between neighbors, and the resistance contribution as a barrier preventing transitions between too dissimilar pixels.

The choice of the dissimilarity D is fairly versatile too: the class of squared Euclidean dissimilarities is broader than often presumed, encompassing L^p dissimilarities for $1 \leq p \leq 2$ (in particular the city-block metric), and all positive semi-definite kernel approaches in machine learning, such as radial basis affinities (e.g. Crichtley and Fichet [14]; Bavaud [6]). D can also express categorical or distributional marks, through chi-square or Hellinger dissimilarities. Covariances between node features (a main theme in Spatial Econometrics; see e.g. Bivand et al. [11]; LeSage [27]; Arbia [3]; Anselin [2]; Griffith [23]) can be taken into account by the use of Mahalanobis dissimilarities. Also, recall that any squared Euclidean dissimilarity D generates, through exact multidimensional scaling, a set of multivariate coordinates X , unique up to a rotation.

Besides its regularizing virtues, the presence of the entropy term in the free energies (6) or (10) can be formally justified in the *maximum a priori* approach of Bayesian statistics (Besag [10] or Greig et al. [22]), the *maximum entropy* approach of Information Theory, or in statistical mechanical models of magnetic materials, where the connection with the Ising or Potts model (in particular regarding the two first “energy” terms of equation (14)) has been often noticed.

Other compatible developments, well-known in spatial analysis or machine learning, such as those involving *Moran scatterplots* and *local indicators of spatial autocorrelation* (Anselin [1]), as well as *spectral approaches* for Ncut or modularity clustering (Shi and Malik [36]; White and Smyth [39]; Ng et al. [31]; von Luxburg [28]) have been left aside. Recall that spectral approaches aim at attacking network clustering by means of a matrix eigen-decomposition problem, sacrificing the non-negative nature of Z , but producing instead “network factor scores” \tilde{X} on which standard clustering algorithms, such as the K-means, can be performed. The latter can be further mixed with node features to form a generalized set of “where-and-what” features (\tilde{X}, X) , on which standard clustering methods can be applied, again; see e.g. Lebichot and Saerens [26] and references therein for a recent presentation. In contrast to our approach, which directly confronts Z to the network E (and its marks X or their dissimilarities D), the latter strategy is closely related to the search for embedding coordinates \tilde{X} for a network, and the definition of associated squared Euclidean *spatial dissimilarities* \tilde{D} (e.g. Yen et al. [40]; Bavaud [5]; Kivimäki et al. [25]), for which the question of the positive definite nature of E (section 5.1) plays a prominent role, again.

References

- [1] Anselin, L.: Local Indicators of Spatial Association – LISA. *Geogr. Anal.* 27(2), 93–115 (apr 1995), <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [2] Anselin, L.: *Spatial econometrics: methods and models*, vol. 4. Springer, Dordrecht (2013)
- [3] Arbia, G.: *Spatial data configuration in statistical analysis of regional economic and related problems*, vol. 14. Springer, Dordrecht (1089)
- [4] Bavaud, F.: Aggregation invariance in general clustering approaches. *Adv. Data. Anal. Classif.* 3(3), 205–225 (nov 2009), <https://doi.org/10.1007/s11634-009-0052-9>
- [5] Bavaud, F.: Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs, pp. 103–118. Springer Berlin Heidelberg (2010), https://doi.org/10.1007/978-3-642-15880-3_13
- [6] Bavaud, F.: On the Schoenberg Transformations in Data Analysis: Theory and Illustrations. *J. Classif.* 28(3), 297–314 (oct 2011), <https://doi.org/10.1007/s00357-011-9092-x>
- [7] Bavaud, F.: Testing spatial autocorrelation in weighted networks: the modes permutation test. *J. Geogr. Syst.* 15(3), 233–247 (jul 2013)
- [8] Bavaud, F., Guex, G.: Random Walks and Shortest Paths: A Path Functional Approach. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) *Social Informatics: 4th International Conference, SocInfo 2012, Lausanne, Switzerland, December 5-7, 2012. Proceedings.* pp. 68–81. Springer Berlin Heidelberg, Berlin, Heidelberg (2012), https://doi.org/10.1007/978-3-642-35386-4_6
- [9] Berger, J., Snell, J.L.: On the concept of equal exchange. *Syst. Res. Behav. Sci.* 2(2), 111–118 (1957), <https://doi.org/10.1002/bs.3830020204>
- [10] Besag, J.: On the Statistical Analysis of Dirty Pictures. *J. R. Stat. Soc. Series B Stat. Methodol.* 48(3), 259–302 (1986), <http://www.jstor.org/stable/2345426>
- [11] Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V., Pebesma, E.J.: *Applied spatial data analysis with R*, vol. 10. Springer, New York, NY (2013)
- [12] Cliff, A.D., O.J.: *Spatial Processes, Models & Applications.* Taylor & Francis (1981)
- [13] Couprie, C., Grady, L., Najman, L., Talbot, H.: Power Watershed: A Unifying Graph-Based Optimization Framework. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(7), 1384–1399 (jul 2011), <https://doi.org/10.1109/TPAMI.2010.200>
- [14] Critchley, F., Fichet, B.: The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties, pp. 5–65. Springer New York (1994), https://doi.org/10.1007/978-1-4612-2686-4_2
- [15] Devooght, R., Mantrach, A., Kivimäki, I., Bersini, H., Jaimes, A., Saerens, M.: Random Walks Based Modularity: Application to Semi-supervised Learning. In: *Proceedings of the 23rd International Conference on World Wide Web.* pp. 213–224. WWW '14, ACM (2014), <https://doi.org/10.1145/2566486.2567986>
- [16] Doyle, P.G., Snell, J.L.: Random walks and electric networks. *Am. Math. Mon.* (1984), <https://doi.org/10.5948/UP09781614440222>
- [17] Fouss, F., Saerens, M., Shimbo, M.: *Algorithms and models for network data and link analysis.* Cambridge University Press (2016), <https://doi.org/10.1017/CB09781316418321>
- [18] Françoisse, K., Kivimäki, I., Mantrach, A., Rossi, F., Saerens, M.: A bag-of-paths framework for network data analysis. *Neural Netw.* 90, 90–111 (jun 2017)

- [19] Gaumer, P., Moliterni, C.: Dictionnaire mondial de la bande dessinée. Larousse (1994)
- [20] Grady, L.: Random Walks for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(11), 1768–1783 (2006), <https://doi.org/10.1109/TPAMI.2006.233>
- [21] Grady, L., Schwartz, E.L.: Isoperimetric graph partitioning for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(3), 469–475 (mar 2006), <https://doi.org/10.1109/TPAMI.2006.57>
- [22] Greig, D.M., Porteous, B.T., Seheult, A.H.: Exact Maximum A Posteriori Estimation for Binary Images. *J. R. Stat. Soc. Series B Stat. Methodol.* 51(2), 271–279 (1989), <http://www.jstor.org/stable/2345609>
- [23] Griffith, D.A.: Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization. Springer, Berlin, Heidelberg (2003)
- [24] Guex, G.: Interpolating between random walks and optimal transportation routes: Flow with multiple sources and targets. *Physica. A* 450, 264–277 (may 2016), <https://doi.org/10.1016/j.physa.2015.12.117>
- [25] Kivimäki, I., Shimbo, M., Saerens, M.: Developments in the theory of randomized shortest paths with a comparison of graph node distances. *Physica. A* 393, 600–616 (jan 2014)
- [26] Lebichot, B., Saerens, M.: An experimental study of graph-based semi-supervised classification with additional node information. *ArXiv* (may 2017)
- [27] LeSage, J.P.: An introduction to spatial econometrics. *Revue d'économie industrielle* (3), 19–44 (2008)
- [28] von Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* 17(4), 395–416 (dec 2007), <https://doi.org/10.1007/s11222-007-9033-z>
- [29] Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* 42(5), 577–685 (jul 1989), <https://doi.org/10.1002/cpa.3160420503>
- [30] Newman, M.E.J.: Modularity and community structure in networks 103(23), 8577–8582 (may 2006)
- [31] Ng, A.Y., Jordan, M.I., Weiss, Y.: On Spectral Clustering: Analysis and an Algorithm. In: *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. pp. 849–856. NIPS'01, MIT Press (2001), <http://dl.acm.org/citation.cfm?id=2980539.2980649>
- [32] Petitot, J.: An introduction to the MumfordShah segmentation model. *J. Physiol. Paris* 97(2), 335–342 (mar 2003), <https://doi.org/10.1016/j.jphysparis.2003.09.007>
- [33] Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* 76, 036106 (sep 2007), <https://doi.org/10.1103/PhysRevE.76.036106>
- [34] Raphaël Céré, a.F.B.: Multi-labelled Image Segmentation in Irregular, Weighted Networks: A Spatial Autocorrelation Approach. In: *Proceedings of the 3rd International Conference on Geographical Information Systems Theory, Applications and Management*. pp. 62–69. Scitepress (2017), <https://doi.org/10.5220/0006322800620069>
- [35] Rose, K., Gurewitz, E., Fox, G.: A deterministic annealing approach to clustering. *Pattern Recognit. Lett.* 11(9), 589–594 (1990)
- [36] Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8), 888–905 (aug 2000), <https://doi.org/10.1109/34.868688>

- [37] Sinop, A.K., Grady, L.: A Seeded Image Segmentation Framework Unifying Graph Cuts And Random Walker Which Yields A New Algorithm. In: 2007 IEEE 11th International Conference on Computer Vision. pp. 1–8 (2007), <https://doi.org/10.1109/ICCV.2007.4408927>
- [38] Vitti, A.: The MumfordShah variational model for image segmentation: An overview of the theory, implementation and use. ISPRS J. Photogramm. Remote Sens. 69, 50–64 (apr 2012), <https://doi.org/10.1016/j.isprsjprs.2012.02.005>
- [39] White, S., Smyth, P.: A Spectral Clustering Approach To Finding Communities in Graphs. In: Proceedings of the 2005 SIAM International Conference on Data Mining (2005)
- [40] Yen, L., Saerens, M., Mantrach, A., Shimbo, M.: A Family of Dissimilarity Measures Between Nodes Generalizing Both the Shortest-path and the Commute-time Distances. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–793. KDD '08, ACM (2008), <https://doi.org/10.1145/1401890.1401984>
- [41] Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation. Tech. rep., Carnegie Mellon University (jun 2002)

7 APPENDIX

7.1 Computing the exchange matrix $E(A, f)$

Defining an exchange matrix E both weight-compatible (that is obeying $E\mathbf{1} = f$, where the regional weights f are given) and reflecting the spatial structure contained in the binary adjacency matrix $A = (a_{ij})$ is a crucial, necessary step in the “ZED formalism” under consideration. Two constructions, not trivial, nor that difficult either, have been investigated in this paper, namely the *diffusive* specification and the *Metropolis-Hastings* specification.

The diffusive exchange matrix

Consider a time-continuous Markov chain W on the n pixels, whose infinitesimal generator or *rate matrix* is proportional to the adjacency matrix A , and conveniently normalized so that f constitutes the stationary distribution of W . The resulting exchange matrix $E = \Pi W$ turns out to be symmetric and p.s.d., and given by

$$E \equiv E(A, f, t) = \Pi^{1/2} \exp(-t\Psi) \Pi^{1/2} \quad (20)$$

where $\Pi = \text{diag}(f)$, and

$$\Psi = \Pi^{-1/2} \frac{LA}{\text{trace}(LA)} \Pi^{-1/2} \quad (LA)_{ij} = \delta_{ij} a_{i\bullet} - a_{ij}$$

LA is the *Laplacian* of matrix A , and matrix exponentiation (20) can be carried out by the spectral decomposition of Ψ . Specification (20) describes a *diffusive process* at *time* $t > 0$, with limits $\lim_{t \rightarrow 0} E(A, f, t) = \Pi$ (“frozen network”, consisting of n isolated nodes: spatial autarchy), and $\lim_{t \rightarrow \infty} E(A, f, t) = ff'$

(“complete network”, with independent selection of the node pairs: complete mobility). Identity $\text{trace}(E(t)) = 1 - t + 0(t^2)$ shows t to measure, for $t \ll 1$, the proportion of *distinct* regional pairs in the joint distribution E .

The Metropolis-Hastings exchange matrix

The natural random walk with Markov transition matrix $a_{ij}/a_{i\bullet}$ correctly describes the spatial structure of the network, but its stationary distribution is $g_i = a_{i\bullet}/a_{\bullet\bullet}$ instead of f_i . Applying the Metropolis-Hastings algorithm defines a recalibrated random walk with stationary distribution f , ending up in a weight-compatible exchange matrix of the form:

$$E = II - LB \quad \text{where} \quad B = (b_{ij}), \quad b_{ij} = \min(\kappa_i, \kappa_j) \cdot \frac{a_{ij}}{a_{\bullet\bullet}} \quad \text{and} \quad \kappa_i = \frac{f_i}{g_i} \quad (21)$$

and $(LB)_{ij} = \delta_{ij} b_{i\bullet} - b_{ij}$ is the Laplacian of B . Expression (21) does not require spectral decomposition, and its computation is much faster than (20) for increasing n (figure 15). However, E in (21) is not p.s.d in general, thus threatening the concavity of $\mathcal{C}^\kappa[Z]$ (section 5.1).

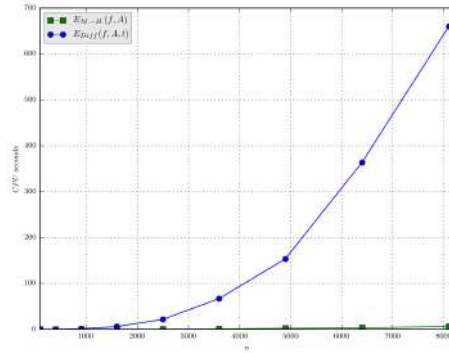


Fig. 15. *Deterministic profiling* : CPU time for computing the exchange matrices $E(f, A, t)$ (20) and $E_{M-H}(f, A)$ (21), as a function of the number of pixels n in a regular setting and performed with *Python* 2.7.12 on a *CPU Intel Core i7* two Core with a frequency 3.1GHz (Mac OS X 10.10.5).

7.2 Testing spatial autocorrelation

Under the null hypothesis H_0 of stationarity and absence of spatial autocorrelation, univariate features are independent, and follow a distribution with common mean and variance *inversely proportional to the size of the region*, namely $E(X_{ik}) = \mu_k$ and $\text{Cov}(X_{ik}, X_{jk}) = \delta_{ij} \sigma_k^2 / f_i$ (Bavaud [7]). Under normal approximation, the expected value of the multivariate Moran’s I (2) reads

$$E_0(I) = \frac{\text{tr}(W) - 1}{n - 1} \quad \text{where} \quad w_{ij} = \frac{e_{ij}}{f_i}$$

and its the variance reads

$$\text{Var}_0(I) = \frac{2}{n^2 - 1} \left[\text{trace}(W^2) - 1 - \frac{(\text{trace}(W) - 1)^2}{n - 1} \right]$$

Spatial autocorrelation is thus significant at level α if $z = |I - E_0(I)| / \sqrt{\text{Var}_0(I)} \geq u_{1-\frac{\alpha}{2}}$, where $u_{1-\frac{\alpha}{2}}$ is the α^{th} quantile of the standard normal distribution.

Alternatively, a permutation test can be performed (e.g. Cliff [12]), by generating a series of values \hat{I} of the transformed Moran index, where \hat{I} obtains as (2) with Δ_{loc} replaced by $\hat{\Delta}_{\text{loc}} = \frac{1}{2} \sum_{i,j=1}^n e_{ij} \hat{D}_{ij}$. The *plain specification*, which consists in replacing the profile x_{ik} of region i by the profile $\hat{x}_{ik} = x_{\pi(i)k}$ of another region $\pi(i)$ (where π denotes a permutation), that is in defining $\hat{D}_{ij} = D_{\pi(i),\pi(j)}$, is somehow flawed in the weighted case, in view of the heteroscedasticity of the distribution of X_{ik} . Instead, the quantities $\sqrt{f_i}(x_{ik} - \bar{x}_k)$ (with $\bar{x}_k = \sum_i f_i x_{ik}$) for $i = 1, \dots, n$ are expected to follow the same distribution under H_0 , thus insuring the validity of the *weight-corrected specification*, with (see figure 2)

$$\hat{x}_{ik} = \bar{x}_k + \sqrt{\frac{f_{\pi(i)}}{f_i}} (x_{\pi(i)k} - \bar{x}_k) \quad \text{and} \quad \hat{D}_{ij} = \|\hat{x}_i - \hat{x}_j\|^2 . \quad (22)$$