Bavaud, F. *Spatial weights: constructing weight-compatible exchange matrices from proximity matrices*. In: Duckham, M. et al. (Eds.): GIScience 2014, LNCS 8728, pp. 81–96. Springer (2014)

Spatial weights: constructing weight-compatible exchange matrices from proximity matrices

François Bavaud

University of Lausanne

Abstract. Exchange matrices represent spatial weights as symmetric probability distributions on pairs of regions, whose margins yield regional weights, generally well-specified and known in most contexts. This contribution proposes a mechanism for constructing exchange matrices, derived from quite general symmetric proximity matrices, in such a way that the margin of the exchange matrix coincides with the regional weights. Exchange matrices generate in turn *diffusive* squared Euclidean dissimilarities, measuring spatial remoteness between pairs of regions. Unweighted and weighted spatial frameworks are reviewed and compared, regarding in particular their impact on permutation and normal tests of spatial autocorrelation. Applications include tests of spatial autocorrelation with diagonal weights, factorial visualization of the network of regions, multivariate generalizations of Moran's I, as well as "*landscape clustering*", aimed at creating regional aggregates both spatially contiguous and endowed with similar features.

Keywords: Euclidean distances, exchange matrix, Laplacian, matrix exponential, Moran's I, permutation test, spatial autocorrelation, spatial coding scheme, spatial weights, `spdep`

1 Introduction

Weighted unoriented networks are specified by node and edge weights. In spatial statistics, node weights f represent the relative importance of regions, normalized to unity, entering into the definition of weighted averages of the form $\bar{x} = \sum_i f_i x_i$. Also, edge weights e_{ij} constitute spatial weights, entering in the definition of spatially autocorrelated models.

Edge weights are *weight-compatible* if their margins coincide with the set of regional weights f, generally well-defined and known a priori. Symmetric, weight-compatible edge weights define an *exchange matrix* E, whose components can be interpreted as *the probability of selecting a pair of regions*.

On one hand, exchange matrices arguably constitute a style of spatial weights particularly adapted to weighted spatial contexts. On the other hand, exchange matrices E are hardly ever directly known to the fellow worker. Instead, the researcher in general only possesses vague, incomplete spatial information, as expressed in a *spatial proximity matrix* G, whose components provide a spatial measure of proximity between pairs of regions. The proximity or generalized adjacency matrix G may represent adjacencies, the size of the common boundary, the inter-regional accessibility, the inter-regional flow of exchanged units (people, matter, goods, information), as well as many other proxies for neighborhood.

Symmetrizing and normalizing G makes it formally equivalent with a distribution on regional pairs, that is with an exchange matrix - see specification U) below. However, the marginal distribution γ resulting from G *does not coincide in general with the regional spatial weights* f: while f measures *regional importance*, γ measures *regional centrality*. Yet, plainly put, a region can be peripheral and important, or central and insignificant, thus establishing the necessity of constructing weight-compatible exchange matrices $E(G, f)$, that is based upon G, but with margin f.

This contribution recalls and reviews a few definitions in spatial autocorrelation (section 2), in unweighted and weighed settings. Particular emphasis is devoted to the comparison of their corresponding canonical measures of spatial autocorrelation, their permutation and normal significance testing, as well as the handling of off-diagonal spatial weights, occurring not that infrequently in applications, such as those involving flows and self-interaction.

The central part (section 3) proposes the construction of a one-parameter family of weight-compatible exchange matrices $E(G, f, t)$ from proximity matrices G. The former, describing a continuous diffusive process generated by G, turns out to be p.s.d., allowing further the definition of a *diffusive squared Euclidean dissimilarity* $\mathcal{D}(G, f, t)$ between regions (section 3.3).

Spatial analysis of French elections illustrate the theory (section 4). Possible applications, briefly outlined in sections (4.1), (4.2) and (4.3), include multivariate generalization of Moran's I, factorial visualization of spatial versus attribute dissimilarities between regions, as well as "*landscape clustering*", aimed at creating regional aggregates both spatially contiguous and endowed with similar characteristics.

2 (Un)weighted measures of spatial autocorrelation

2.1 Unweighted setting: spatial weights from spatial links $V(G)$

In presence of n regions of equal importance (uniform weighting) characterized by the density variable x, Moran's index of spatial autocorrelation is usually defined as (e.g. Cliff and Ord 1981; Anselin 1995; Tiefelsdorf and Boots 1995; Dray et al. 2006)

$$I \equiv I(V, x) := \frac{n \sum_{ij} v_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{ij} v_{ij}) \sum_i (x_i - \bar{x})^2} \qquad \bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where the *spatial weights* matrix $V = (v_{ij})$ is non-negative, symmetric or not.

By construction, $I \equiv I(V, x)$ does depend upon on the spatial field x under investigation, as well as, crucially, upon the specification of spatial weights V:

see e.g. Tiefelsdorf et al. (1999) for an explicit illustration. The latter authors also propose and investigate various *spatial coding schemes* aimed at extracting a convenient spatial weights matrix $V(G)$ from *spatial link* or *proximity* matrices $G = (g_{ij})$, meant as an immediate, possibly rough but accessible spatial information about proximity relationships between regions i and j. Proximities G between regions may be determined by mutual contiguity, accessibility, inverse distance, flow, etc. In what follows, we assume G to be *symmetric* $g_{ij} = g_{ji}$ as well as *essentially non-negative*, that is such that $g_{ij} \geq 0$ for $i \neq j$. Typical choices are

i) $g_{ij} = a_{ij}$: binary adjacency or contiguity matrix
ii) $g_{ij} = n_{ij}$, where n_{ij} counts the number of units (people, matter, money, information) flowing from i to j
iii) $g_{ij} = F(d_{ij})$ where d_{ij} is a measure of the distance between i and j and $F(d) \geq 0$ is a distance-deterring, decreasing function
iv) $g_{ij} = |\partial A_{ij}|$, the measure of the common boundary between distinct regions i and j.

In particular, Tiefelsdorf et al. (1999) together with other workers have considered the following *coding schemes* $V(G)$:

B) the *binary spatial weights* $v_{ij} = 1(g_{ij} > 0)$ taking on value one if $g_{ij} > 0$, and zero otherwise
W) the *row-standardized spatial weights* $v_{ij} := g_{ij}/g_{i\bullet}$ (where $\bullet$ denotes summation over the replaced index, as in $g_{i\bullet} := \sum_j g_{ij}$), prevalent in models of spatial autocorrelation
C) the *globally standardized spatial weights* $v_{ij} := n g_{ij}/g_{\bullet\bullet}$
U) the *standardized spatial weights* $v_{ij} := g_{ij}/g_{\bullet\bullet}$
S) the *variance-stabilizing spatial weights*

$$v_{ij} := \frac{n\, s^*_{ij}}{\sum_{ij} s^*_{ij}} \qquad \text{where} \qquad s^*_{ij} := \frac{g_{ij}}{\sqrt{\sum_j g^2_{ij}}} \ .$$

The above spatial coding schemes respectively constitute the so-called B, W, C, U and S schemes, as referred to and used in the `spdep` R package (Bivand 2002; Bivand et al. 2006).

2.2 Weighted setting: E-coding scheme $E(G, f, t)$

In all generality, the importance of the n regions differ, as quantified by their relative weights $f_i > 0$ with $\sum_{i=1}^n f_i = 1$. Typically, regional weights f_i reflect the relative population (human geography), relative area (physical geography) or relative wealth (economic geography) of region i. Regional spatial weights f_i can be interpreted as the probability $P(i)$ of selecting region i.

Specifying a symmetric probability $P(i,j) = e_{ij}$ to select a pair of neighboring regions (i,j) defines the *exchange matrix* $E = (e_{ij})$ (Berger and Snell 1957). By construction,

$$e_{ij} = e_{ji} \geq 0 \qquad e_{i\bullet} = f_i > 0 \qquad \sum_{ij} e_{ij} = e_{\bullet\bullet} = 1 \ .$$

In this weighted setup, Moran's index of spatial autocorrelation reads (e.g. Bavaud 2013 and references therein)

$$I \equiv I(E,x) := \frac{\sum_{ij} e_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i f_i(x_i - \bar{x})^2} \qquad \bar{x} := \sum_{i=1}^{n} f_i x_i \tag{2}$$

In particular, $-1 \leq I(E,x) \leq 1$ with expected value $E_0(I(E,x)) = -1/(n-1)$ under absence of spatial autocorrelation, *provided E contains no diagonal components* (see section 6 for the general case). Note the equivalent formulation

$$I(E,x) = \frac{\mathrm{var}(x) - \mathrm{var}_{\mathrm{loc}}(x)}{\mathrm{var}(x)} \tag{3}$$

where

$$\mathrm{var}(x) = \frac{1}{2}\sum_{ij} f_i f_j (x_i - x_j)^2 = \sum_i f_i (x_i - \bar{x})^2$$

is the ordinary (weighted) variance and

$$\mathrm{var}_{\mathrm{loc}}(x) = \frac{1}{2}\sum_{ij} e_{ij}(x_i - x_j)^2$$

is the *local variance*, measuring the average dissimilarity between pairs of spatially associated regions (e.g. Lebart 1969; Le Foll 1982; Meot et al. 1993; Thioulouse et al. 1995). The concept of local variance is related, yet distinct, to the concept of *local indicator of spatial autocorrelation* (Anselin 1995), referring to a weighted decomposition of Moran's I (3) as in $I(E,x) = \sum_i f_i I_i(E,x)$.

The row-standardized matrix of spatial weights $W = (w_{ij})$ obtains from the exchange matrix as $w_{ij} = e_{ij}/f_i$, and constitutes the transition matrix of a reversible Markov chain (Bavaud 1998).

3 Obtaining the exchange matrix in two steps (4) and (5)

Given a symmetric and essentially non-negative, "off-positive" proximity matrix G, i.e. whose off-diagonal components are non-negative, as well as a set of regional weights f, compute the symmetric matrix $\Psi = (\psi_{ij})$

$$\psi_{ij}(G,f) = \frac{1}{\sqrt{f_i f_j}}\,\frac{\delta_{ij} g_{i\bullet} - g_{ij}}{(g_{\bullet\bullet} - \mathrm{trace}(G))} \ . \tag{4}$$

Then compute the exchange matrix by means of the matrix exponential

$$E(t) := \Pi^{1/2} \exp(-t\,\Psi)\Pi^{1/2} \qquad \text{where} \qquad \Pi = \mathrm{diag}(f) \quad \text{and} \quad t \geq 0 \ . \tag{5}$$

The free parameter $t > 0$ interprets as the *age of the network*. By construction (proofs below):

1) $E(t)$ in (5) is symmetric and weight compatible: $e_{i\bullet}(t) = f_i$ for any $t \geq 0$
2) $e_{ij}(t) \geq 0$ for all i, j and $t \geq 0$
3) $E(t)$ is p.s.d. (section 3.2)
4) $\lim_{t\to 0} e_{ij}(t) = \delta_{ij} f_i$, expressing complete regional segregation in a "frozen network"
5) $\lim_{t\to\infty} e_{ij}(t) = f_i f_j$, which expresses absence of distance-deterrence effects in a "free" or "complete network".

3.1 Further formal considerations

The numerator in (4) contains the so-called *Laplacian* of G (e.g. Chung 1997 p.12), defined as $(LG)_{ij} := \delta_{ij} g_{i\bullet} - g_{ij}$, where δ_{ij} are the components of the identity matrix (Kronecker's delta). By construction, LG is *positive semi-definite* (p.s.d.) (see section 3.2) and obeys

$$\mathrm{trace}(LG) = \mathrm{sum}(G) - \mathrm{trace}(G) \ \geq \ 0 \qquad\qquad \mathrm{sum}(LG) = 0 \tag{6}$$

where $\mathrm{sum}(C) := c_{\bullet\bullet}$ and $\mathrm{trace}(C) := \sum_i c_{ii}$. In particular, and $L\,\mathrm{diag}(b) = 0$ *for any diagonal matrix* $\mathrm{diag}(b)$ with diagonal b: hence the diagonal elements a_{ii} of the adjacency matrix in (i) (loops), the stayers flow n_{ii} in (ii), or self-boundaries $|\partial A_{ii}|$ in (iv) play *no role* in the construction of LG, Ψ or $E(t)$. $\Psi(G)$ is indeed invariant with respect to transformations of the form $G \to a(G + \mathrm{diag}(b))$, for any scalar $a > 0$ and any vector b (cf. (6) and (4)).

Normalizing $\Psi(G)$ as in (4) amounts in normalizing t in (5), in the hope of making the scale t "intrinsic" or "absolute", that is hopefully comparable among differing data sets. As a matter of fact,

$$\Psi = \Pi^{-1/2}\,\frac{LG}{\mathrm{trace}(LG)}\,\Pi^{-1/2} \qquad\qquad E(t) = \Pi \ - \ \frac{LG}{\mathrm{trace}(LG)}\,t \ + \ O(t^2) \tag{7}$$

The question of the choice of t itself remains fairly open so far. The "self-exchange proportion" $\mathrm{trace}(E(t))$ decreases with t, converging to $\mathrm{trace}(E(\infty)) = \sum_i f_i^2 < 1$, with small time expansion $\mathrm{trace}(E(t)) = 1 - t + 0(t^2)$. This proportion could possibly be estimated by $\mathrm{trace}(N)/\mathrm{sum}(N)$, where N is the inter-regional flows matrix, or some other measure of spatial interaction. For instance, inter-cantonal migrations between 1985 and 1990 in Switzerland yields $1 - \hat{t} = \mathrm{trace}(N)/\mathrm{sum}(N) = 0.93$ (most people stayed in the same canton during those five years), yielding the possible estimate $\hat{t} = 0.07$.

Equations (4) and (5) constitute a straightforward two-steps procedure generalizing and simplifying the "proposal B" recipe exposed in Bavaud (2013), based upon the construction of a weight-compatible, time-continuous Markov chain generated by a *rate matrix* $R = -\Pi^{-1/2}\Psi\Pi^{1/2} = -\Pi^{-1}LG/\text{trace}(LG)$ reflecting direct spatial transitions, as expressed by the proximity matrix G, whose regional sojourn times are precisely adjusted to make $E(t)$ weight-compatible.

Non-negativity condition 2) above is a direct consequence of the essential non-positivity of Ψ together with the theorem "$\exp(-t\,A)$ *is non-negative for all* $t > 0$ *iff* A *is essentially non-positive*" (see e.g. theorem 8.2 in Varga 2000).

3.2 Spectral decomposition

Solution (5) can be computed by spectral decomposition of $\Psi = UMU'$, that is $\psi_{ij} = \sum_\alpha \mu_\alpha u_{i\alpha} u_{j\alpha}$. As a matter of fact, $\Psi\sqrt{f} = 0$, thus $\sqrt{f}$ is a trivial eigenvector u (numbered $\alpha = 0$) of Ψ with trivial eigenvalue $\mu_0 = 0$, demonstrating in particular the weight-compatibility

$$E(t)\mathbf{1} = \Pi^{1/2}\sum_{r=0}^{\infty}\frac{(-t)^r}{r!}\Psi^r\sqrt{f} = \Pi^{1/2}\sqrt{f} = f$$

as claimed. The other eigenvalues, *increasingly ordered*, are non-negative, since Ψ turns out to be p.s.d. in view of

$$0 \le \frac{1}{2}\sum_{ij} g_{ij}(h_i - h_j)^2 = \sum_{ij}(\delta_{ij}g_{i\bullet} - g_{ij})h_i h_j$$

for any h. Thus $0 = \mu_0 \le \mu_1 \le \mu_2 \le \ldots \le \mu_{n-1}$ and

$$E(t) = \Pi^{1/2}\exp(-tUMU')\Pi^{1/2} = \Pi^{1/2}U\exp(-Mt)U'\Pi^{1/2}$$

that is

$$e_{ij}(t) = \sqrt{f_i f_j}\sum_{\alpha=0}^{n-1} u_{i\alpha}u_{j\alpha}\exp(-\mu_\alpha t) = f_i f_j + \sqrt{f_i f_j}\sum_{\alpha=1}^{n-1} u_{i\alpha}u_{j\alpha}\exp(-\mu_\alpha t)$$

thus proving limits 4) and 5) above. Equivalently,

$$e^s_{ij}(t) := \frac{e_{ij} - f_i f_j}{\sqrt{f_i f_j}} = \sum_{\alpha=1}^{n-1} u_{i\alpha}u_{j\alpha}\exp(-\mu_\alpha t) \qquad E^s = U\exp(-Mt)U' \qquad (8)$$

where $E^s(t) := \Pi^{-1/2}(E(t) - ff')\Pi^{-1/2}$ is the *standardised exchange marix.* $E^s(t)$ possesses a trivial eigenvalue $\lambda_0 = 0$ associated with $u_0 = \sqrt{f}$, as well as non-trivial eigenvalues $\lambda_\alpha(t) = \exp(-\mu_\alpha t)$, *decreasingly ordered* for $\alpha \ge 1$, lying in $[-1, 1]$, as required by the Perron-Froebenius theorem on the associated Markov chain W. They even lie in $[0, 1]$, making $E(t)$ p.s.d. or *diffusive.* Note the eigenvectors $U = (u_{i\alpha})$ to be independent of t.

3.3 Diffusive dissimilarity and multidimensional scaling

The p.s.d. nature of $E(t)$ permits to define a "diffusive dissimilarity" between regions, namely

$$\mathcal{D}_{ij}(t) := \frac{e_{ii}(t)}{f_i^2} + \frac{e_{jj}(t)}{f_j^2} - 2\,\frac{e_{ij}(t)}{f_i f_j} \quad . \tag{9}$$

$\mathcal{D}$ turns out to be *squared Euclidean*, i.e. of the form $\mathcal{D}_{ij} = \|y_i - y_j\|^2$ for some $n \times p$ "diffusive coordinates" $Y = (y_{ia})$, where $p \leq n-1$. The squared Euclidean nature of $\mathcal{D}$ follows from the *conditional negative-definiteness condition* $\sum_{ij} h_i h_j \mathcal{D}_{ij} \leq 0$ for any h with $\sum_i h_i = 0$ (see e.g. Cressie 1993). Determining the diffusive coordinates $Y = (y_{i\alpha})$ constitutes the classical multidimensional scaling (MDS) problem, with solutions in the weighted setting

$$y_{i\alpha}(t) = \frac{\sqrt{\lambda_\alpha(t)}}{\sqrt{f_i}}\, u_{i\alpha} \qquad\qquad \alpha = 1, \ldots, n-1 \tag{10}$$

where $\lambda_\alpha(t) = \exp(-\mu_\alpha t)$ is the eigenvalue of E^s in (8), and $u_{i\alpha}$ its corresponding eigenvector (e.g. Cuadras and Fortina 1996; Bavaud 2010). (10) is a member of a family of vertex coordinates on weighted graphs of the form

$$y_{i\alpha}(t) = g(\lambda_\alpha(t))\, y^s_{i\alpha} \qquad\qquad y^s_{i\alpha} = \frac{u_{i\alpha}}{\sqrt{f_i}} \qquad\qquad \alpha = 1, \ldots, n-1 \tag{11}$$

where $g(\lambda)$ is a non-negative function, and $y^s_{i\alpha}$ is the *standardized* or *raw coordinate* of region i on dimension $\alpha \geq 1$ (Bavaud 2010). Raw coordinates also occur quite naturally in spatial filtering (e.g. Griffith 2000, 2003; Dray et al. 2006; Bavaud 2013).

3.4 Summary

Any proximity matrix G between regions, together with any set of regional weights f, yield a one-parameter family of weight-compatible, p.s.d. exchange matrices $E(G, f, t)$. The latter yield in turn a family of squared Euclidean dissimilarities $\mathcal{D}_{ij}(t)$ between regions (9), from which regional coordinates $y_{i\alpha}(t)$ (10) or raw coordinates $y^s_{i\alpha}$ (11) can be extracted by weighted MDS. Hence, any pair (G, f) produces a visualization y or y^s of the spatial configuration between regions, conceivably resembling the true geographical configuration (Figure 1).

4 Illustration: political autocorrelation in France

Consider the $n = 94$ departments of "metropolitan France" (Corsica excluded), whose binary adjacency matrix A is chosen as the spatial matrix. Consider also uniform departmental weights $f_i = 1/n$, but also, in parallel, non-uniform "voters weights" f (section 4.1). Figure 2 depicts the distribution of departmental degrees $a_{i\bullet}$ and non-uniform weights f_i, as well as the non-trivial eigenvalues $\lambda_\alpha(t)$. Figure 1 gives the first two factorial "raw coordinates" (11), in the uniform and non-uniform case. The reconstruction of the geographical map from the adjacency matrix looks fairly adequate.

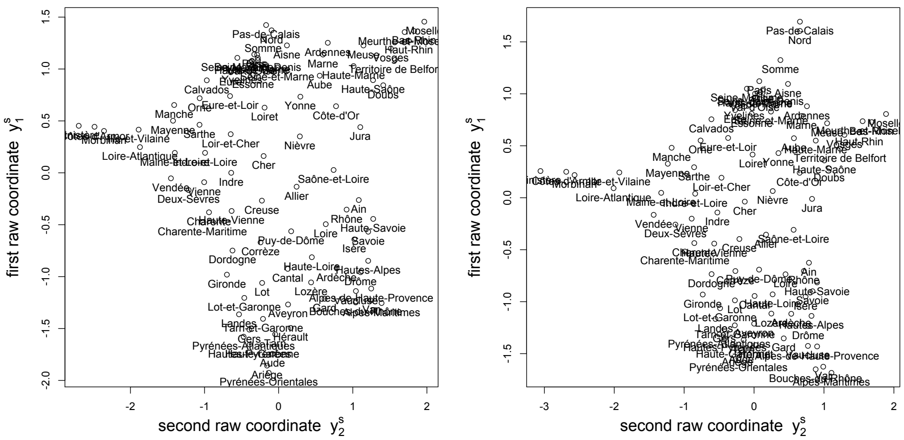

Fig. 1. Raw diffusive coordinates $y^s_{i\alpha}$ (11) for $\alpha = 2$ (abscissa) and $\alpha = 1$ (ordinate), reconstructing the map of French departments from the adjacency matrix A and departmental weights f. Left: uniform weights $f_i = 1/n$. Right: non-uniform "voters weights".

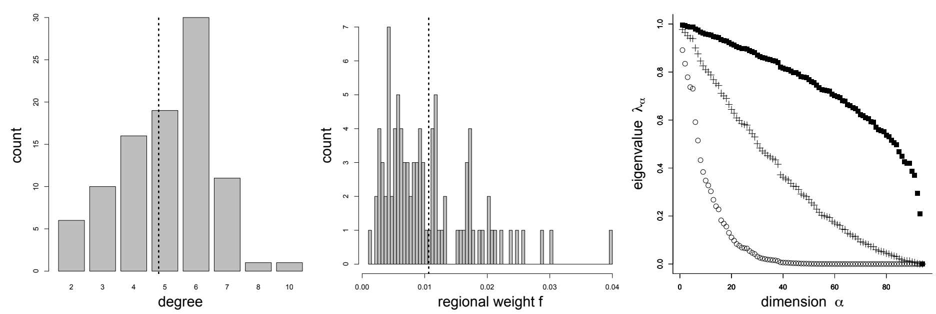

Fig. 2. Left: distribution of the departmental degrees $a_{i\bullet}$ (average degree = 5.06). Middle: distribution of the non-uniform departmental "voters weights" f_i (section 4.1; average weight = 1/94=0.011). Right: scree plot of the eigenvalues $\lambda_\alpha(t)$ of $E^s(t)$ (non-uniform weights) for $t = 0.2$ (solid squares), $t = 1$ (crosses) and $t = 5$ (circles).

4.1 Extracting regional features by Correspondence Analysis

In general, regions are characterized by uni- or multivariate *features* x, whose variation may or may not be correlated with the diffusive coordinates y in (10); this issue precisely constitutes the topic of *spatial autocorrelation*, as exemplified below. In the sequel, features x will first be computed as regional factor scores, instead of considering directly available regional variables x.

Consider the votes of the first round of the French presidential 2012 election, as recorded by the $n \times p$ contingency table (N_{ik}), fixing the "number of votes for candidate k in department i" where $n = 94$ and $k = 1, \ldots, p = 10$ (Joly, LePen, Sarkozy, Melenchon, Poutou, Arthaud, Cheminade, Bayrou, Dupont-Aignan, Hollande). Figure 3 left yields the scree plot of the proportion of explained chi-square by each of the $\min(n, p-1) = 9$ factors, whose first and second ones express together 83% of the inertia. Figure 3 right exposes the Correspondence Analysis (CA) biplot depicting the *department* and *candidate coordinates*, showing similarities among departments, among candidates, as well as attraction-repulsion between departments and candidates.

In this context, natural regional weights are provided by $f_i = N_{i\bullet}/N_{\bullet\bullet}$, the *voters weight* of department i, measuring its relative share of voters. By construction, department coordinates $x_{i\beta}$ are centered, standardized and uncorrelated (here $\beta = 1, \ldots p-1$ labels the factors produced in Correspondence Analysis):

$$\sum_i f_i x_{i\beta} = 0 \qquad \sum_i f_i x_{i\beta}^2 = 1 \qquad \sum_i f_i x_{i\beta} x_{i\beta'} = 0 \quad \text{for } \beta \neq \beta' \ .$$

4.2 Spatial autocorrelation of voting pattern

The autocorrelation of each "voting factor" x_β (where $\beta = 1, \ldots, p$), as extracted from the CA of section (4.1), can be tested in turn by computing Moran's indices $I(E, x_\beta)$ in (2). Here $E = E(A, f, t)$ is the weight-compatible, time-dependent exchange matrix constructed in section (3.3), and f stands as before as the non-uniform voters weight.

Associated p-values are computed from B bootstrapped samples associated with the weighted permutation test (section 6). The first factorial political score x_{i1} extracted in section (4.1) turns out to be strongly autocorrelated (Figure 4, left), in contrast to the second score x_{i2} which is not (Figure 4, right).

4.3 Relative inertia

Moran's I can be generalized to *multivariate settings* by considering squared Euclidean dissimilarities D_{ij} between regional profiles, instead of univariate dissimilarities of the form $(x_i - x_j)^2$. In the present analysis, the natural candidate for D is provided by the classical *chi-square dissimilarity* between departments,

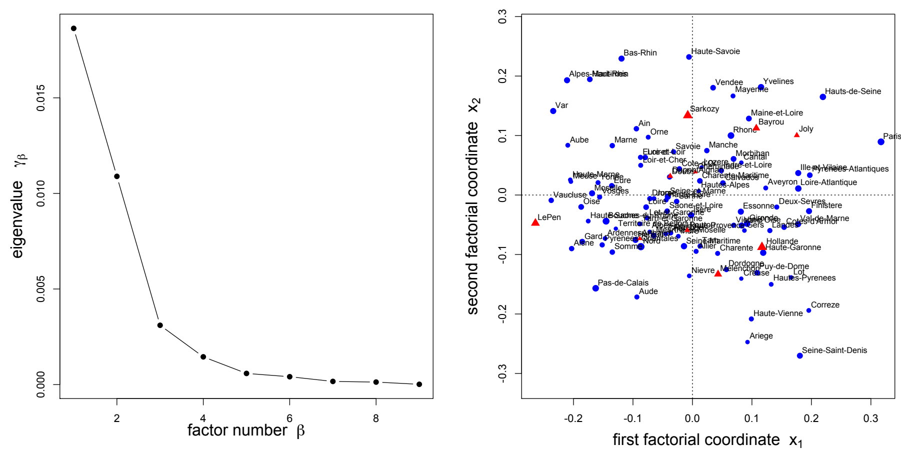

Fig. 3. Correspondence Analysis on the "department × candidate" votes contingency table, in the first round of the French presidential 2012 election. Left: eigenvalues γ_β. Right: biplot. The first axis can be interpreted in terms of right- versus left-wing contrast, but also central-peripheral contrast (53%); the second one seems to oppose poor to rich departments (31%).

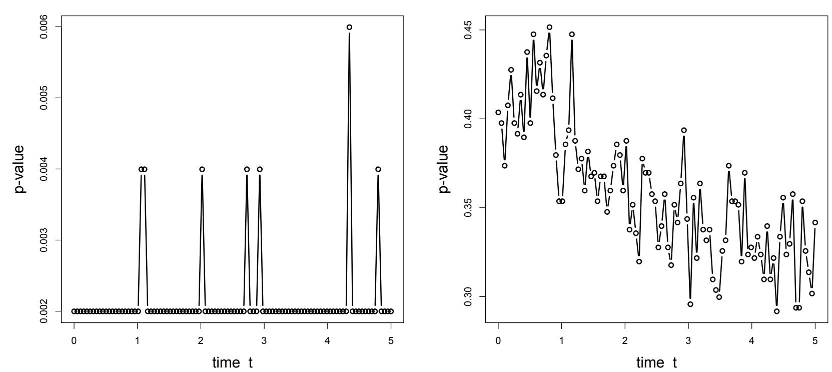

Fig. 4. p-values associated to the significance test of weighted Moran's I (3), for various values of $t \in [0, 5]$, based upon $B = 1'000$ permutations of the so-called *spatial modes* (instead of the direct spatial values), in order to take into consideration the heteroscedasticity associated to the weighted setting (see section 6 and Bavaud 2013 for details). The first political component x_{i1} turns out to be strongly autocorrelated (left), in contrast to the second political component x_{i2} (right).

which can be defined from the contingency table (N_{ik}) or from the "raw" factor scores $x_{i\beta}$, as

$$D^{\chi}_{ij} = \sum_k \frac{N_{\bullet\bullet}}{N_{\bullet k}}(\frac{N_{ik}}{N_{i\bullet}} - \frac{N_{jk}}{N_{j\bullet}})^2 = \sum_\beta \gamma_\beta (x_{i\beta} - x_{j\beta})^2 \qquad (12)$$

where γ_β are the eigenvalues (the square of the singular values) of the Correspondence Analysis of section 4.1. Recall (and observe) that weighted multidimensional scaling of D precisely yields the so-called *principal coordinates* $\sqrt{\gamma_\beta}\, x_{i\beta}$, that is *CA is equivalent to weighted MDS on chi-square dissimilarities*.

As claimed, multivariate generalization of Moran's index (3) is provided by the *relative inertia* $\delta \in [-1, 1]$ defined (with $D = D^\chi$) as

$$\delta(t) := \frac{\Delta - \Delta_{\text{loc}}(t)}{\Delta} \qquad \Delta = \frac{1}{2}\sum_{ij} f_i f_j D_{ij} \qquad \Delta_{\text{loc}}(t) = \frac{1}{2}\sum_{ij} e_{ij}(t) D_{ij}$$

whose significance can be assessed by usual normal approximation, permutation, or bootstrap tests; see section 6.2. Relative inertia also expresses as (cf. (1))

$$\delta(t) = \frac{\sum_{ij} e_{ij}(t)\, B_{ij}}{\sum_i f_i B_{ii}}$$

where B are the scalar products of MDS, obeying $D_{ij} = B_{ii} + B_{jj} - 2B_{ij}$. In particular, $B_{ii} = D_{if}$ is the squared distance between i and the centroid.

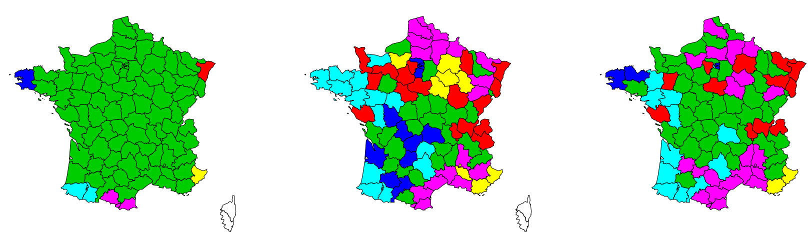

Fig. 5. Soft K-means clustering of French departments, from the initial centroid configuration determined by the $m = 6$ "Hexagon corners": spatial clustering (left), attributes clustering (middle) and landscape clustering (right) defined in (13), with $\nu = 1/2$ and $c(D) = \max_{ij} \mathcal{D}_{ij}$.

4.4 "Landscape clustering"

We are now in possession of *two squared Euclidean dissimilarities*, namely the diffusive dissimilarity $\mathcal{D}_{ij}$ (9), measuring *spatial remoteness* between pairs of regions, and the chi-square dissimilarity D^{χ}_{ij} (12), measuring the voting *attributes* *contrast* between regional profiles. This circumstance makes it possible to consider various *regional clustering strategies*, namely

a) *spatial clustering*, based upon the diffusive dissimilarity $\mathcal{D}_{ij}$ exhibited in Figure 2 right
b) *attributes clustering*, based upon the attributes dissimilarity D^{χ}_{ij} exhibited in Figure 3 right
c) a presumably new *landscape clustering* based upon minimizing the within-groups inertia associated to the mixed, normalized dissimilarity

$$D^{lan}_{ij} = \nu \, \frac{\mathcal{D}_{ij}}{c(\mathcal{D})} + (1-\nu) \, \frac{D^{\chi}_{ij}}{c(D^{\chi})} \tag{13}$$

where $\nu \in (0,1)$ controls the spatial versus attribute contributions and $c(D)$ is a normalisation factor, such as $\frac{1}{2}\sum_{ij} f_i f_j D_{ij}$ or $\max_{ij} \mathcal{D}_{ij}$. Landscape clustering aims at creating regional aggregates both spatially contiguous and possessing similar features - a natural aim in Quantitative Geography, Spatial Econometrics and Geographic Information Science.

Figures 5 and 6 illustrate spatial clustering (left), attributes clustering (middle) and landscape clustering (right) for the mixed normalised dissimilarities Clusterings result from soft K-means (section 6.3), with initial centroid configuration determined by the $m = 6$ "Hexagon corners" (Bas-Rhin, Nord, Finistère, Pyrénées-Atlantiques, Pyrénées-Orientales, Alpes-Maritimes: Figure 5), or by the $m = 7$ most populated departments (Nord, Bouches-du-Rhône, Paris, Rhône, Pas-de-Calais, Gironde, Loire-Atlantique: Figure 6).

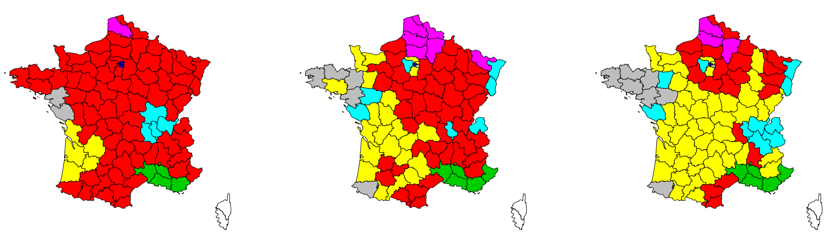

Fig. 6. Soft K-means clustering of French departments, from the initial centroid configuration determined by the $m = 7$ most populated departments: spatial clustering (left), attributes clustering (middle) and landscape clustering (right) defined in (13), with $\nu = 1/2$ and $c(D) = \max_{ij} \mathcal{D}_{ij}$.

5 Discussion and Conclusions

Dealing with regions of unequal importance requires a weighted formalism, which arguably helps unifying mathematical enquiries and proposals. For instance, Moran's I and Geary's c appear to be simply related as $c = 1 - I$ in the present "E-scheme".

After briefly reviewing the differences between the unweighted and weighted approaches to spatial autocorrelation, this paper proposes a straight, general prescription aimed at constructing exchange matrices E both compatible with given proximity relations G and regional weights f. The solution contains a freely adjustable parameter t, the age of the network, controlling the importance of direct adjacency, distance deterrence, or inverse bandwidth, when $0 < t < \infty$. At the extremes, the network becomes independent of G, namely with the frozen network $E(0) = \Pi$ and the completely mobile network $E(\infty) = ff'$.

Solution $E(t)$ turns out to be p.s.d., that is modeling a diffusive network. This circumstance permits to define a squared Euclidean dissimilarity on the network, and hence, by MDS, a network visualisation. This presumably new *proximity-based dissimilarity* can in turn be compared to some other *features-based* squared Euclidean dissimilarity: this constitutes the very issue of spatial autocorrelation. Both similarities can also be mixed, and fed to partitionning algorithms, yielding "landscape clustering", sensitive to both regional proximities and attributes.

More case studies are most welcome. Further investigations could examine the impact of $E(G, f, t)$ on weighted SAR or CAR models, on the construction of mobility indices, or on the construction of *local indicators* of relative inertia, in the spirit of the well-known proposal of Anselin (1995).

6 Appendix: autocorrelation tests and soft clustering

The first part of the appendix derives, under the null hypothesis H_0 of no autocorrelation, the expected value of Moran's I and its variance in the general case of spatial weights, *possibly containing non-zero diagonal components* $v_{ii} \neq 0$, a case little confronted with in the literature. Both unweighted (section 6.1) and weighted settings (section 6.2) are addressed.

6.1 Unweighted permutation test

Equation (1) shows that V can be taken as **symmetric** and **normalized**, that is obeying $v_{ij} = v_{ji}$ and $v_{\bullet\bullet} = 1$. Moran index thus expresses as (e.g. Griffith 2003; Dray et al. 2006)

$$I(V, x) = \frac{n \sum_{ij} v_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} = n\frac{x'HVHx}{x'Hx} \qquad\qquad H = I - J/n \ .$$

Here I is the identity matrix, $J = \mathbf{11'}$ is the constant unit matrix and $H = H^2$ is the centering projection matrix, each of order $n \times n$.

The spectral decomposition $HVH = U\Lambda U'$ with U orthogonal and Λ diagonal makes appear a trivial dimension $\alpha = 0$, with constant eigenvector $u_{i0} = 1/\sqrt{n}$ and null eigenvalue $\lambda_0 = 0$. Also, $Hu_\alpha = u_\alpha$ for higher-order, non-trivial dimensions $\alpha = 1, \dots n-1$.

Define the unweighted *spatial modes* (see e.g. Griffith 2003 and Bavaud 2013) as $y = U'x$, that is $x = Uy$. In particular, $y_0 = \sqrt{n}\,\bar{x}$. Moran's index then reads

$$I(V,x) = n\frac{\sum_{\alpha\geq 1}\lambda_\alpha y_\alpha^2}{\sum_{\alpha\geq 1} y_\alpha^2} = n\sum_{\alpha\geq 1}\lambda_\alpha b_\alpha \qquad b_\alpha := \frac{y_\alpha^2}{\sum_{\beta\geq 1} y_\beta^2} \tag{14}$$

In the present unweighted setting, the spatial variables X_i are, under H_0, i.i.d. with mean μ and variance σ^2. The resulting spatial modes $Y = U'X$ are uncorrelated with $E(Y_\alpha) = \delta_{\alpha 0}\sqrt{n}\mu$ and $\mathrm{Cov}(Y_\alpha, Y_\beta) = \delta_{\alpha\beta}\sigma^2$. In particular, the $n-1$ quantities b_α $(\alpha \geq 1)$ in (14) are arguably identically distributed under H_0, yet not independently in view of $\sum_{\alpha\geq 1} b_\alpha = 1$. Denoting by $E_\pi(.)$ the expectation under modes permutation, one gets, by symmetry

$$E_\pi(b_\alpha) = \frac{1}{n-1} \quad E_\pi(b_\alpha^2) = \frac{\sum_{\beta\geq 1} b_\beta^2}{(n-1)} =: \frac{t(y)}{(n-1)^2} \quad E_\pi(b_\alpha b_\beta) = \frac{1-t(y)/(n-1)}{(n-1)(n-2)}$$

for $\alpha \neq \beta$. Taking into account

$$\sum_{\alpha\geq 1}\lambda_\alpha = \sum_{\alpha\geq 0}\lambda_\alpha = \mathrm{trace}(HVH) = \mathrm{trace}(V) - \frac{1}{n} \qquad \text{and}$$

$$\sum_{\alpha\geq 1}\lambda_\alpha^2 = \sum_{\alpha\geq 0}\lambda_\alpha^2 = \mathrm{trace}(HVHHVH) = \mathrm{trace}(VHVH) = \mathrm{trace}(V^2) - \frac{2}{n}\sum_i v_{i\bullet}^2 + \frac{1}{n^2}$$

finally yields

$$E_\pi(I) = \frac{n\,\mathrm{trace}(V) - 1}{n-1} \qquad \boxed{\text{unweighted setting}}$$

$$\mathrm{Var}_\pi(I) = \frac{t(y)-1}{(n-1)(n-2)}[n^2\mathrm{trace}(V^2) - 2n\sum_i v_{i\bullet}^2 + 1 - \frac{(n\,\mathrm{trace}(V)-1)^2}{n-1}] \tag{15}$$

where $t(y) = (n-1)\sum_{\alpha\geq 1} y_\alpha^4/(\sum_{\alpha\geq 1} y_\alpha^2)^2 \geq 1$ is a measure of modes dispersion, taking on its minimum value $t(y) = 1$ for $y_\alpha = \mathrm{const}$ for $\alpha \geq 1$. In particular, $E_\pi(I) > -1/(n-1)$ whenever spatial weights V contain off-diagonal components.

Under the additional normal assumption $X_i \sim N(\mu, \sigma^2)$, one gets $Y_\alpha \sim N(0, \sigma^2)$ for $\alpha \geq 1$, as well as $E(t(y)) = 3(n-1)/(n+1)$ (e.g. Cliff and Ord 1983, p.43). Then

$$\begin{aligned} E(\mathrm{Var}_\pi(I)) &= \frac{2}{n^2-1}[n^2\mathrm{trace}(V^2) - 2n\sum_i v_{i\bullet}^2 + 1 - \frac{(n\,\mathrm{trace}(V)-1)^2}{n-1}] \qquad (16) \\ &= \frac{1}{(n^2-1)S_0^2}[n^2 S_1 - nS_2 + \frac{2(n-2)S_0^2 + 4nS_0\mathrm{trace}(V) - 2n^2\mathrm{trace}^2(V)}{n-1}] \end{aligned}$$

where, for comparison's sake, the normalization condition $v_{\bullet\bullet} = 1$ has been relaxed in the last equation (while retaining the symmetry of V), and the familiar notations

$$S_0 := \sum_{ij} v_{ij} \qquad S_1 := 2\sum_{ij} v_{ij}^2 \qquad S_2 := 4\sum_{i} v_{i\bullet}^2$$

have been introduced. The last identity in (16) turns out to coincide, up to the terms involving $\mathrm{trace}(V)$, with the formulas proposed in Cliff and Ord (1981, p.44).

6.2 Weighted permutation test

In the weighted setup, the spatial field X_i represents a regional aggregate associated to region i. Under H_0, its mean is constant but its variance is inversely proportional to the weight of the region (heteroscedasticity). Hence, $X_i \sim N(\mu, \sigma^2/f_i)$ under normal assumption. The expected value of the weighted Moran coefficient (2) and its variance read (Bavaud 2013)

$$E_\pi(I) = \frac{\mathrm{trace}(W) - 1}{n-1} \qquad \boxed{\text{weighted setting}}$$

$$E(\mathrm{Var}_\pi(I)) = \frac{2}{n^2-1}[\mathrm{trace}(W^2) - 1 - \frac{(\mathrm{trace}(W)-1)^2}{n-1}] \qquad (17)$$

where $W = (w_{ij})$ with $w_{ij} := e_{ij}/f_i$ is the row-normalized, weight-compatible matrix of spatial weights (section 2.2), and constitutes the transition matrix of a Markov chain, possessing off-diagonal components in general (Bavaud 1998).

Unweighted average and variance formulas (15) and (16) should coincide with their weighted analogs (17) whenever $V = E$ constitutes a symmetric, normalized spatial weight matrix with uniform weights $e_{i\bullet} = f_i = 1/n$. Indeed, $W = nV$ with $v_{i\bullet} = 1/n$ in that case, thus demonstrating the expected agreement.

Small time limit In the limit $t \to 0$, (7) and (17) together with $W(t) = I + tR + \frac{t^2}{2}R^2 + 0(t^2)$, where $R = -\Pi^{-1}LG/\mathrm{trace}(LG)$ is the rate matrix, yield

$$I - E_\pi(I) = \frac{t}{\mathrm{trace}(LG)}[\frac{\mathrm{trace}(\Pi^{-1}LG)}{n-1} - \frac{\sum_{ij} g_{ij}(x_i - x_j)^2}{2\,\mathrm{var}(x)}] + 0(t^2)$$

$$E(\mathrm{Var}_\pi(I)) = \frac{2t^2}{n^2-1}[\mathrm{trace}(R^2) - \frac{\mathrm{trace}^2(R)}{n-1}] + 0(t^3) \qquad \boxed{\text{weighted setting, } t \to 0}$$

Interestingly enough, for uniform weights $f_i = 1/n$, the decision variable of the normal test expresses, up to order $0(t)$, as

$$z = \frac{I - E_\pi(I)}{\sqrt{E(\mathrm{Var}_\pi(I))}} = \frac{\tilde{I} - E_\pi(\tilde{I})}{\sqrt{E(\mathrm{Var}_\pi(\tilde{I}))}}$$

where

$$\tilde{I}(x) := 1 - \frac{1}{\mathrm{var}(x)} \frac{\frac{1}{2}\sum_{ij} g_{ij}(x_i - x_j)^2}{g_{\bullet\bullet} - \mathrm{trace}(G)}$$

is *time-independent*, and constitutes an alternative to unweighted Moran's index $I(x)$ (1).

6.3 Soft K-means

Consider a $n \times m$ *membership matrix* Z with components $z_{ig} \geq 0$ with $z_{i\bullet} = 1$, expressing the probability that region i belongs to group g. The group weight is $\rho_g[Z] = \sum_i f_i z_{ig}$ and the squared Euclidean dissimilarity D_i^g between region i and centroid g is derived form the inter-individual squared dissimilarities D_{ij} as (Huygens principle)

$$D_i^g[Z] = \sum_j f_j^g D_{ij} - \frac{1}{2}\sum_{ij} f_i^g f_j^g D_{ij} \qquad \text{where} \quad f_i^g[Z] = \frac{f_i z_{ig}}{\rho_g} \ .$$

where D_{ij} is a squared Euclidean dissimilarity. Memberships are iteratively computed (e.g. Celeux and Govaert 1992, Rose 1998, Bavaud 2010) as

$$z_{ig}^{(r+1)} = \frac{\rho_g[Z^{(r)}] \ \exp(-\beta D_i^g[Z^{(r)}])}{\sum_h \rho_h[Z^{(r)}] \ \exp(-\beta D_i^h[Z^{(r)}])}$$

where the *inverse temperature* β has been set to 1 in section 4.4, where two variants for the m initial centroids are investigated. After convergence, region i is finally attributed to group $g = \arg\max_h z_{ih}^{(\infty)}$. The alternative iteration

$$f_i^{g(r+1)} = \frac{f_i \ \exp(-\beta D_i^g[Z^{(r)}])}{\sum_j f_j \ \exp(-\beta D_j^g[Z^{(r)}])}$$

works as well, as expected.

References

Anselin, L., 1995, "Local indicators of spatial association – LISA", Geographical analysis 27, pp. 93–115

Bavaud, F., 1998, "Models for Spatial Weights: A Systematic Look", Geographical Analysis 30, pp. 153–171

Bavaud, F., 2010, "Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs". In Balcázar et al. (Eds), "ECML PKDD 2010", Lecture Notes in Artificial Intelligence 6321, pp. 103–118 (Springer)

Bavaud, F., 2013, "Testing spatial autocorrelation in weighted networks: the modes permutation test", Journal of Geographical Systems 15, pp. 233–247

Berger, J., Snell, J.L., 1957, "On the concept of equal exchange", Behavioral Science 2, pp. 111–18

Bivand, R., 2002, "Spatial econometrics functions in R: Classes and methods". Journal of Geographical Systems 4, pp. 405–421

Bivand, R., Anselin, L., Berke, O., Bernat, A., Carvalho, M., Chun, Y., ... & Yu, D., 2006, “spdep: Spatial dependence: weighting schemes, statistics and models” [R package]

Celeux G., Govaert G., 1992 “A classification EM algorithm and two stochastic versions”, Computational Statistics and Data Analysis 14, pp. 315–332

Chung, F.R.K., 1997, “Spectral Graph Theory”, AMS

Cliff, A. D., Ord, J. K., 1981, “Spatial processes”, Pion

Cressie, N., 1993, “Statistics for Spatial Data”, Wiley

Cuadras, C.M., and Fortina, J., 1996, “Weighted continuous metric scaling”, In: A.K. Gupta and V.L. Girko (Eds.), Multidimensional Statistical Analysis and Theory of Random Matrices, VSP, the Netherlands, 27–40

Dray, S., Legendre, P., Peres-Neto, P.R., 2006, “Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)”, Ecological Modelling, Volume 196, pp. 483–493

Griffith, D.A., 2000, “Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses”, Linear Algebra and its Applications, Vol.321, pp. 95112

Griffith, D.A., 2003, “Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization”, Springer

Lebart, L., 1969, “Analyse Statistique de la Contiguïté”, Publications de l’Institut de Statistique des Universités de Paris, XVIII, pp. 81–112

Le Foll, Y., 1982, “Pondération des distances en analyse factorielle”, Statistique et Analyse des Données, 7, pp 13–31

von Luxburg, U., 2007, “A tutorial on spectral clustering”, Statistics and Computing 17, pp. 395–416

Meot, A., Chessel, D., Sabatier, R. 1993, “Opérateurs de voisinage et analyse des données spatio-temporelles”. In : D. Lebreton and B. Asselain (Eds.), Biométrie et Environnement, Masson, Paris, pp. 45–71

Rose, K. 1998, “Deterministic Annealing for clustering, compression, classification, regression, and related optimization problems”, Proceedings of the IEEE 86, pp. 2210–2239

Saporta, G. 1990, “Probabilités, analyse des données et statistique”, Editions Technip

Thioulouse, J., Chessel, D., Champely, S. 1995, “Multivariate analysis of spatial patterns - a unified approach to local and global structures”, Environmental and Ecological Statistics 2, pp. 1–14

Tiefelsdorf, M., Boots, B., 1995, “The exact distribution of Moran’s I”, Environment and Planning A 27, pp. 985–999

Tiefelsdorf, M., Griffith, D. A., Boots, B., 1999, “A variance-stabilizing coding scheme for spatial link matrices”, Environment and Planning A 31, pp. 165–180

Varga, R. S., 2000, “Matrix Iterative Analysis”, Springer