



HAL
open science

Using landscape graphs to delineate ecologically functional areas

Jean-Christophe Foltête, Gilles Vuidel

► **To cite this version:**

Jean-Christophe Foltête, Gilles Vuidel. Using landscape graphs to delineate ecologically functional areas. *Landscape Ecology*, 2017, 32 (2), pp.249-263. 10.1007/s10980-016-0445-z . hal-01443809

HAL Id: hal-01443809

<https://hal.science/hal-01443809>

Submitted on 4 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Jean-Christophe Foltête*, Gilles Vuidel

Théma, UMR 6049 CNRS – Université Bourgogne-Franche-Comté, 32, rue Mégevand F-25030 Besançon, France

*Corresponding author

This is a postprint version, the definitive version of this paper is :

Foltête, J.-C., & Vuidel, G. (2017). Using landscape graphs to delineate ecologically functional areas. *Landscape Ecology*, 32(2), 249-263. <https://doi.org/10.1007/s10980-016-0445-z>

Abstract

Context

Landscape graphs are widely used to model connectivity and to support decision-making in conservation planning. Compartmentalization methods applied to such graphs aim to define clusters of highly interconnected patches. Recent studies show that compartmentalization based on modularity is suitable, but it applies to non-weighted graphs whereas most landscape graphs involve weighted nodes and links.

Objectives

We propose to adapt modularity computation to weighted landscape graphs and to validate the relevance of the resulting compartments using demographic or genetic data about the patches.

Methods

A weighted adjacency matrix was designed to express potential fluxes, associating patch capacities and inter-patch distances. Eight weighting scenarios were compared. The statistical evaluation of each compartmentalization was based on Wilks' Lambda. These methods were performed on a grassland network where patches are documented by annual densities of water voles in the Jura massif (France).

Results

The scenarios in which patch capacity is assigned a small weight led to the more relevant results, giving high modularity values and low Wilks' Lambda values. When considering a fixed number of compartments, we found a significant negative correlation between these two criteria. Comparison showed that compartments are ecologically more valid than graph components.

Conclusions

The method proposed is suitable for designing ecologically functional areas from weighted landscape graphs. Maximum modularity values can serve as a guide for setting the parameters of the adjacency matrix.

Keywords

Landscape connectivity, compartmentalization, modularity, ecological network, montane water vole

1. Introduction

Preserving biodiversity is a major concern in the face of climate change and anthropogenic land cover changes. It implies preserving the habitats of wildlife species and maintaining their accessibility for individuals at different spatial scales, so as to ensure populations remain viable. In the context of heterogeneous landscapes where habitat takes the form of a set of discrete patches, landscape connectivity is a key notion for scientists and land managers dealing with conservation issues (Taylor et al. 2006).

Among the different methods designed to represent habitat networks and provide decision support tools for managers, landscape graphs (or habitat graphs) have been used increasingly for the last fifteen years. These methods strike a valuable compromise between their capacity to model potential for functional connectivity and the limited data they require as input (Minor and Urban 2007; Urban et al. 2009). The general principle of these methods is to consider habitat patches as the nodes of the graph and the inter-patch connections as the links (Urban and Keitt 2001). Beyond this general definition, different types of landscape graphs may be designed depending on the specific nature of each case study, the purpose of the analysis, and the technical resources the user can mobilize (Galpern et al. 2011; Urban et al. 2009). Landscape graphs may be used first to provide ecological knowledge (or at least assumptions that can potentially be validated in the field), about for example the population model underlying the graph configuration (Urban et al., 2009) or the effect of habitat connectivity on species distribution (Awade et al. 2011; Foltête et al. 2012a; Lookingbill et al. 2010). From another perspective, their purpose is to provide a framework for facilitating environmental assessment and decision-making in conservation and land planning, by offering the users the opportunity to compare multiple alternatives. The primary contributions these methods may make from a land planning perspective are in (1) identifying the most vulnerable areas of an ecological network in which to prioritize conservation measures, (2) locating the areas in which to implement field actions (e.g. landscaping) so as to improve overall connectivity, (3) assessing and mapping the impact on species of a change in land cover (Foltête et al., 2014).

As mentioned in Gao et al. (2013), few studies have focused on the notion of compartments in landscape graphs, although this is a fundamental notion in other applications of graph theory such as social networks (Girvan and Newman 2002). Since compartments are defined as groups of nodes within which functional relationships are stronger than with other nodes, they may be highly relevant in terms of wildlife population persistence (Minor and Urban 2007; Urban et al. 2009) and may provide a suitable way for delineating functionally homogeneous areas. Landscape graph compartments may also be included in landscape genetics approaches pertaining to the boundary method (Hall and Beissinger 2014) and compared with spatial clusters defined from genetic data. Moreover, since these compartments are expected to have specific demographic dynamics, they may be used to adapt the design of population monitoring plans. Graph compartments are therefore suitable objects for addressing many questions about ecological networks and landscape connectivity. How, then, should such compartments be designed?

Several approaches are available for generating landscape graph compartments. Bodin and Nordberg (2007) experimented with the method proposed by Girvan and Newman (2002) for defining compartments in a patch-based graph. This method is based on the betweenness centrality index applied to the links, i.e. the theoretical proportion of transit across each link resulting from the shortest paths joining every pair of patches. The algorithm of Girvan and Newman (2002) is a top-down (i.e. divisive) process consisting in successively removing the links with the highest betweenness values. Adopting an agglomerative strategy, Gao et al. (2013) used the modularity criterion proposed by Newman and Girvan (2004a) and compared the resulting compartments with those obtained using the previous method based on betweenness. The authors concluded that the modularity-based method yields compartments characterized by suitable properties in terms of internal connectivity, compartment size, and network traversability. This method seems suitable for generating relevant compartments in a landscape graph. However, the case study investigated in Gao et al. (2013) has certain specific features that could limit the generalization of this conclusion to other patch-based

graphs. The graph analyzed by those authors was thresholded according to the dispersal distance of the target species and was composed of uniformly weighted patches and links. This type of graph (i.e. non-weighted graph) is common in many application domains of graph theory and is suitable for applying the method of Newman and Girvan (2004a). However, it is not representative of the patch-based graphs usually designed in landscape ecology (Galpern et al. 2011) that often include two other features to better represent biological processes:

(1) In most landscape graphs, patches are associated with a capacity attribute representing their demographic potential (Urban and Keitt 2001). This allows the patches analyzed to be differentiated by the amount of potential resources they offer (Minor and Urban, 2007). Patch area is the usual criterion considered as a proxy for demographic potential, but other criteria may be considered, such as the terrestrial habitat of amphibians (Clauzel et al. 2015) or foraging areas in the case of point nodes (Tournant et al. 2013).

(2) Next, when landscape matrix heterogeneity is supposed to influence the movement behavior of the target species, the inter-patch links must be characterized by an “effective distance”, most often defined from edge-to-edge least-cost paths. In this case, the links are weighted by the total cost of the shortest path across the matrix.

Since graphs based on area-weighted patches and cost-weighted links are commonplace in landscape ecology, it has to be asked whether the clustering methods based on modularity are fit for purpose, and if they are not, what could be done to ensure better matching with weighted graphs.

Another issue with the definition of compartments in a landscape graph is the ecological validation of the resulting compartmentalization. A first evaluation level consists in investigating the spatial pattern of the results, for example by checking whether the networks within the compartments have certain expected topological properties, as in Bodin and Norberg (2007) and Gao et al. (2013). A second level of validation would be to check the relevance of the compartments by comparing them in terms of their demographic or genetic characteristics. Data describing demographic variations could be used to validate graph compartments especially with species subject to strong population fluctuations and for which functionally connected areas may give rise to asynchronous dynamics between compartments. With regard to genetics, as compartments represent closely inter-connected groups of habitat patches, they may be the source of some genetic differentiation if their spatial configuration remains stable during a sufficient time. Consequently, comparing genetic clusters with graph compartments would be a suitable method of validation, in the perspective proposed by Luque et al. (2012). So far as we know, no experiment has yet been conducted to validate graph compartments using genetic or demographic data. The main reason for this is probably the difficulty of systematically acquiring field data across a large area on which to carry out a statistical analysis in a regional-level graph. Such an approach could nevertheless be useful for testing the theoretical assumptions about the biological flows underlying landscape graphs.

In the paper, we propose to adapt modularity computation to the weighted graphs commonly met in landscape ecology, so as to include patch capacity and link weighting in the process of generating graph compartments. We hypothesize that global modularity computed from a weighted graph is a reliable indicator of well-delineated, ecologically functional areas. To test this hypothesis by using multivariate data about species (e.g. demographic variables) and by comparing several scenarios of graph weighting, a statistical protocol based on Wilks’ Lambda is presented. We illustrate the method by a case study focused on cyclic fluctuations of populations of the montane water vole, a grassland rodent spreading in the Jura massif (eastern France), for which a large amount of demographic data have been collected.

2. Methods

2.1. Defining compartments using modularity

Modularity is a criterion for evaluating the quality of a partition of a network and for defining community structures. It was initially proposed by Newman and Girvan (2004a) in the field of social networks. It was recently presented in landscape ecology in the study of Gao et al. (2013) and applied

to population graphs in Cavanaugh et al. (2014). Given a graph and a partition of its nodes into subsets, called compartments, modularity is based on comparison of the connections actually observed and the theoretical connections of a random partition. With a graph containing n nodes and m links, modularity M of a given partition P is defined following Newman (2004b) as:

$$M(P) = \frac{1}{2m} \sum_{i=1}^n \sum_{j=1}^n \left(A_{ij} - \frac{d_i d_j}{2m} \right) \gamma_{ij}$$

where A_{ij} is the adjacency matrix, d_i and d_j are the respective degrees of the nodes i and j , and γ_{ij} is a matrix representing the compartments such that:

$$\gamma_{ij} = \begin{cases} 1 & \text{if the nodes } i \text{ and } j \text{ belong to the same compartment of } P \\ 0 & \text{otherwise} \end{cases}$$

In this formula, A_{ij} represents the observed connections and $\frac{d_i d_j}{2m}$ represents the theoretical connections for a random partition. It should be noted that the degree d_i is computed from the matrix A such that $d_i = \sum_{j=1}^n A_{ij}$. As a result, modularity is close to 1 when most of the links are located inside the compartments, close to -1 otherwise, and close to 0 if the links are randomly distributed.

Let us compare the expression of modularity with different types of landscape graph commonly used. Following the previous formula, the computation of modularity rests on a binary structure expressed by the matrix A , where two given nodes i and j are adjacent ($A_{ij}=1$) or not ($A_{ij}=0$). In the case of a landscape graph, it is assumed we have a non-weighted graph where the links represent a connection that is considered probable ($A_{ij}=1$), such as those for which the distance is below a given threshold, the others being omitted ($A_{ij}=0$). However, as mentioned in the introduction, patch capacity and link weighting are often included in landscape graphs, because these criteria make the graphs more realistic (Urban et al. 2009). With regard to the links, transforming distances into binary adjacencies in the form of thresholded graphs is a suitable procedure for generating a visual representation of the ecological network, but at the same time it may cause significant loss of information (Moilanen 2011). Likewise, considering all the patches to be uniformly weighted leads to mix large patches with high demographic potential that act as sources of dispersal and small patches subject only to immigration from other patches or crossed as stepping stones (Urban and Keitt 2001).

The inclusion of patch capacity and link weighting in the model is thus useful for quantifying the potential interactions between habitat patches. The computation of modularity can be adapted to these continuous criteria by replacing the adjacency matrix A with a weighted adjacency matrix W (Newman et al. 2004b; Barrat et al. 2004):

$$M(P) = \frac{1}{2m} \sum_{i=1}^n \sum_{j=1}^n \left(W_{ij} - \frac{d_i d_j}{2m} \right) \gamma_{ij}$$

In the case of a landscape graph, the values of the weighted adjacency matrix must represent a potential flux between patches. Consequently, we propose to define W such that:

$$W_{ij} = (a_i a_j)^\beta e^{-\alpha d_{ij}}$$

where a_i and a_j are the capacity of patches i and j , d_{ij} is the distance between these patches (Euclidean distance or least-cost distance), β is a parameter controlling the relative weight of the patch capacities, and α is a parameter representing the effect of distance on the dispersal probability. The use of the weighted adjacency matrix induces a change in the term d_i in the modularity formula that is now a weighted degree (Barrat et al. 2004) such that:

$$d_i = \sum_{j=1}^n W_{ij}$$

Likewise the number of links m in the non-weighted graph becomes the total sum of the weights (divided by two because the links are symmetric):

$$m = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n W_{ij}$$

Given all these changes from the initial definition, modularity can be used to evaluate the partition of a landscape graph with variable patch capacity and link weighting.

Using modularity to evaluate the quality of a weighted graph partition, there is now the matter of defining the partition involving the maximum modularity value. However, it is impossible to find the optimal partition in the general case (Brandes et al. 2008) and a strategy must be chosen to find an approximate solution.

The approach commonly used to define compartments with high levels of modularity is based on greedy agglomeration (Brandes et al. 2008). The greedy algorithm starts from singletons (i.e. each patch is in an isolated compartment) and iteratively merges two compartments that yield a partition providing the largest increase (or the smallest decrease) in modularity. When all the compartments have been merged, the partition with the highest modularity is selected. Brandes et al. (2008) have shown from several examples that the greedy algorithm performs well and deviates from optimum modularity by 10% at most in the cases studied by these authors. The resulting partition can be improved by a local optimization. Here the vertex mover algorithm (VM) (Schuetz and Caflich 2008) has been added to optimize modularity after applying the greedy algorithm. This procedure moves nodes between neighboring compartments from one cluster to another if the move increases modularity. The process stops when no more moves can be found that increase modularity. Interestingly, the greedy and VM algorithms work for both weighted and non-weighted graphs, therefore they can be applied to all types of landscape graphs.

2.2. Landscape graph design in view of compartmentalization

An important issue in graph compartmentalization is the type of graph the user wants to submit to this procedure. As described in Galpern et al. (2011) and Foltête et al. (2012b), many choices have to be made when building these graphs. Apart from patch capacity and link weighting, the global geometry of a landscape graph likely to be submitted to a clustering process must be questioned here. Two main options of geometry are available when building a landscape graph: a complete topology, in which all inter-patch links are possible, and a planar topology restricting the links to the nearby patches and avoiding two links overlapping.

The complete topology involves generating a thresholded graph, i.e. selecting the links up to a given maximum distance; otherwise, $n(n-1)$ links would have to be managed with n patches. In this case, as in the studies of Bodin and Norberg (2007) and Gao et al. (2013), thresholding may create several graph components whereas the clustering process must be applied in a single component. Following these authors, the analysis may focus on the main component, but in certain contexts where habitat is too fragmented, the thresholded graph should not be a suitable base for applying compartmentalization. Planar topology involves choosing a way of defining the links. To do this, the minimal planar graph (MPG) (Fall et al. 2007) based on the Voronoi polygons around patches is frequently used. As the MPG corresponds to a single component, it is an appropriate framework for defining compartments while including a weighting of patches and links. Consequently, this type of graph will be investigated in this paper. It should be noted that the MPG should be less suitable for the clustering based on modularity without the weighting of patches and links, because in this case, the node degree depends only on the local density of nodes and does not properly reflect the strength of interactions between nodes.

2.3. Validation of a partition with field data

The validation approach requires a set of data pertaining to the target species in the habitat patches. A large proportion of the habitat patches must be documented by these data to make a statistical analysis possible. In the following, we will present a generic framework by noting x_1, x_2, \dots, x_p the p variables characterizing the patches ($p > 1$), that may represent various types of field data particular to each case study. They may be for example genetic measurements, densities or abundances, synthetic demographic indicators of species, acquired at a single date or over several time periods. Whatever the nature of these data, the statistical analysis will be relevant if it is assumed that close and strongly connected patches have more common characteristics than distant and weakly connected patches. Given a set of n documented patches, a partition will be considered as relevant if the compartments present an internal variance smaller than the global variance. This well-known principle formalized by the Huygens theorem can be generalized to multivariate data, for example in the framework of the linear discriminant analysis or in multivariate analysis of variance (McLachlan 2004). Among the different criteria available for evaluating the quality of a statistical partition, Wilks' Lambda is commonly used (Everitt and Dunn 1991; McLachlan 2004). Wilks' Lambda is based on the ratio between within-classes (here within-compartments) variance and total variance, as follows:

$$L = \frac{|W|}{|T|}$$

where $|W|$ is the determinant of the within-compartments variance-covariance matrix and $|T|$ is the determinant of the total variance-covariance matrix. Wilks' Lambda ranges from 0 when the patches belonging to the same compartment are identical for all the variables (i.e. perfect partition), to 1 when the patches belonging to the same compartment differ as much as with the other patches (i.e. meaning an irrelevant partition). This criterion may be combined with a chi-square test using the approximation proposed by Barlett (Everitt and Dunn 1991). It is worth noting that computation of L may be biased by strong collinearities between the variables x .

Wilks' Lambda can serve to compare several partitions with the same number of compartments, but cannot be directly used to compare nested partitions included in a hierarchical clustering. This is because the within-compartments variance automatically decreases as the number of compartments increases. However, as the modularity-based clustering follows an agglomerative process here, the identification of the suitable number of compartments must rely on comparisons drawn from a quality indicator. In this perspective, an interesting property of L depending on the number of compartments g has been outlined by Marriott (1971) and experimented by Webster (1972). These authors have shown that, in the condition of uniform or unimodal distribution of the variables x , L decreases with a factor g^2 when k increases by a unit. Consequently, g^2L can be used as a quality indicator of the partition, to find the optimal number of compartments (minimal value of g^2L) and to identify several key levels in the nested partition, given the shape of the curve of g^2L according to g . Moreover, a partition will be specifically relevant when $g^2L < 1$ (Webster, 1972).

Considering that a landscape graph may be a suitable representation of the ecological network of a given species, the graph compartments are expected to be characterized by a quite homogeneous ecological function. Consequently, we suppose that for a series of partitions generated from the same graph, modularity M , Wilks' Lambda L , and the indicator g^2L are linked by statistical relationships where L and g^2L should decrease as M increases.

3. A case study: partition of the grassland network in the plateau of Jura (France)

3.1. Study area and context

The study area extends over 5000 km² of the Jura massif, on the border between France and Switzerland. Its western part is a series of karst plateaus rising from 400 m in the west to 1000 m in the east, bounded by deep valley gorges and mostly covered by grassland and forest. In the eastern part, the Folded Jura is characterized by higher relief culminating at 1700 m and landscapes dominated by evergreen coniferous forests.

In this area, populations of montane water vole (*Arvicola terrestris sherman*) spread in grasslands over several decades as cyclic outbreaks, acting as a travelling wave with five-to-eight-year cycles (Blant et al. 2009; Giraudoux et al. 1997). Many studies have been carried out to identify the determinants of this phenomenon, showing the effect of landscape composition, landscape configuration, and agricultural practices on the diffusion of water vole populations. These studies have shown that water voles “swarm” more intensively and more rapidly in homogeneous grassland openfields than in heterogeneous mosaics of grasslands, hedgerows, and wooded patches (Berthier et al. 2009; Duhamel et al. 2000; Morilhat et al. 2008; Berthier et al., 2014). In this context, the purpose is to use a landscape graph to delineate areas subjected to the same demographic dynamics, to provide a support for defining a long-term plan for monitoring water vole populations.

3.2. Land cover and demographic data

Land cover was mapped by combining three databases. Buildings, the hydrographic network, transport infrastructures, and forests (deciduous, evergreen and mixed) were extracted from the French land-cover database (BD Topo IGN 2010). Agricultural areas classified into two categories (annual crops and grassland) were taken from a farming database (BD Agreste 2010). Ponds and wetlands were taken from a special-purpose database (BD Zones Humides DREAL). In addition, the main hedgerows were identified by applying a morphological analysis (MSPA, see Vogt et al., 2007) on the basis of the layer describing forests. Ultimately, a map containing 10 land-cover classes was produced in raster format with a spatial resolution of 10 m (Fig. 1).

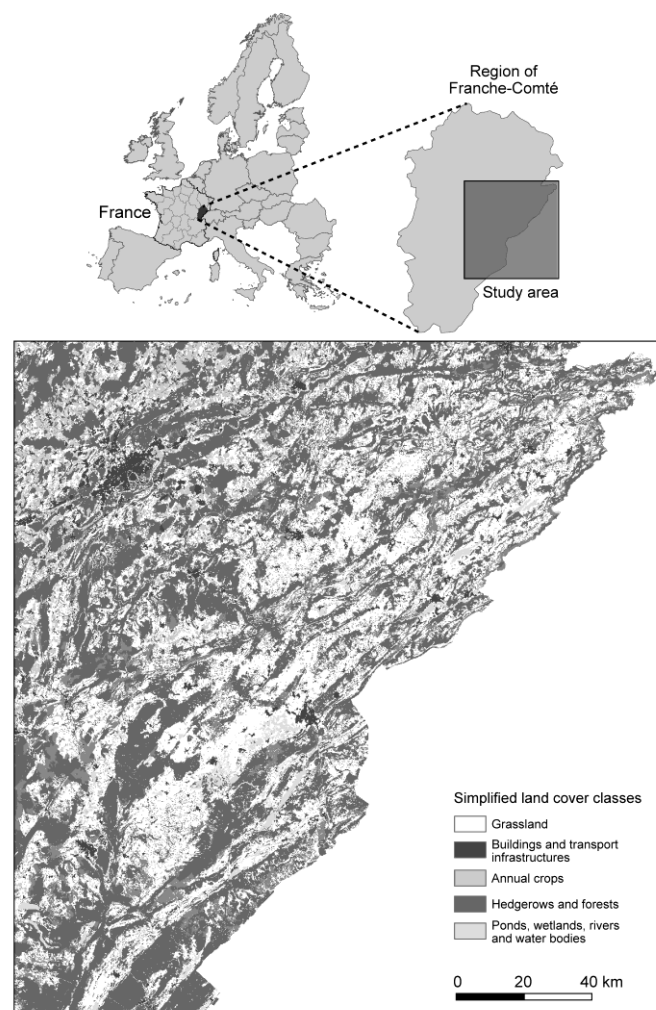


Figure 1 : Land cover of the study area. The main town (Besançon) is located in the north-west part of this area. Except from this urban area, landscape is dominated by forests and agricultural land. The annual crops are mainly located in low areas (northwest) whereas the proportion of grasslands increases with altitude towards the southeast.

The graph nodes were the compact grassland patches, defined as areas of grassland resulting from 50 m erosion followed by 50 m expansion and generating two subclasses: patches (compact) and other (elongate) grassland areas. To limit the number of features to be analyzed, only patches measuring at least 5 ha were included. By this protocol 4132 patches were identified.

Water vole population densities have been estimated for every year from 1989 to 2004 for 312 administrative districts in the French part of the Jura. Field observations have been used to evaluate population density according to an ordinal score (Giraudoux et al. 1995) in which 0 corresponds to fewer than 50 individuals per hectare; 1 to about 50 individuals per hectare; 2 to about 150, and 3 to 400 and more. These data cover three cycles, the first from 1990 to 1994, the second from 1995 to 1999, and the third from 2000 to 2004.

These data were integrated in the graph as in Foltête and Giraudoux (2012) and Foltête et al. (2016). In the central part of the study area involved in population monitoring since 1989 (1905 grassland patches), known annual water vole population densities at the administrative district scale were attributed to grassland patches, with weighting by relative surface areas for patches astride more than one administrative district. From these values, the first-differenced time series of log-densities were calculated following Bjornstadt et al. (1999). As a result of this operation, each grassland patch was characterized by a series of 14 inter-annual variations of water vole density.

3.3. Graph settings

The links of the minimal planar graph were defined as the least-cost paths between patches and associated with the cumulative cost. To define these paths, cost values assigned to land cover classes were provided by the study of Foltête and Giraudoux (2012) in which several cost scenarios were statistically compared to two synthetic indicators derived from the demographic data about water voles. The scenario providing the best correlation between the cost distances and the demographic indicators was used in the present paper and is presented in Table 1. In Foltête and Giraudoux (2012), this maximum correlation was found when a distance of 1000 cost units (approximately 4 km) was matched with a probability of movement of 0.5. This setting was used in the present analysis.

Table 1. Cost values assigned to the land cover classes for the computation of least-cost distances.

Land cover classes	Cost value
Patches of grassland	1
Other grasslands	1
Built environment	50
Transport networks	50
Crops	25
Hedgerows	25
Broad-leaf forest	1000
Mixed forest	1000
Conifer forest	50
Wetlands	50
Streams and ponds	1000

Values defined following Foltête and Girardot

Several weighted adjacency matrices were submitted to the clustering process, each resulting from a specific setting. The setting concerned the definition of the parameter β controlling the relative role of the patch areas in the weighting of the matrix W (see section 2.1.), the other parameter α being fixed in such way that $e^{-\alpha d} = 0.5$ for $d = 1000$. Eight scenarios noted S were compared by setting β to the following values: 0 (S1), 0.1 (S2), 0.25 (S3), 0.5 (S4), 0.75 (S5), 1 (S6), 1.25 (S7), 1.5 (S8).

In scenario S1 where β was set to 0, only the distance associated with the links was included in the weighting, meaning $W = e^{-\alpha d}$. In the following scenarios, growing importance was attributed to the patch areas in the weighting of W according to the constant weight assigned to the distances.

3.3. Results

Graphs were constructed, modularity-based compartments generated, and Wilks' Lambda computed using Graphab 2.0 (Foltête et al., 2012b, see <http://thema.univ-fcomte.fr/productions/graphab/>).

The first step of the analysis was to investigate the capacity of the different weighted adjacency matrices to provide clusters associated with high levels of modularity. For each scenario of weighted adjacency matrix, graph clustering was performed by applying the greedy algorithm seeking to maximize modularity for an increasing number of compartments. Fig. 2a shows the resulting values of modularity according to the number of compartments ranging from 2 to 1905. Globally, all curves exhibit the same behavior, with a rapid increase between 2 and about 10 compartments (modularity > 0.5 for all scenarios) and then a decrease as the number of compartments increases, leading to the minimum values of modularity (< 0.5) with 1905 compartments whatever the scenario. Certain scenarios, specifically 2 and 3, generate peak-shaped curves while others (6, 7, and 8) produce plateau-shaped curves.

We focus now on the number of compartments, ranging from 2 to 20, because the calculation of Wilks' Lambda requires at least two patches per compartment and this constraint is no longer fulfilled for certain scenarios when more than 20 compartments are generated. Fig. 2b allows us to focus on the beginning of the modularity curves. The shapes of these curves are globally similar, but the scenarios clearly differ by the maximum level of modularity, between 0.6 and 0.9. The scenario 1 (where flux between patches is defined as a function of distance alone) has a medium value of around 0.7. The other scenarios display maximum modularity values ordered in the opposite direction to the weight given to the patch areas, from scenario 2 reaching a maximum modularity of 0.87 to scenario 8 reaching only a value of 0.63. In addition, the curves of scenarios 2 to 4, achieving the highest values of modularity, stabilize for a greater number of compartments than the other curves.

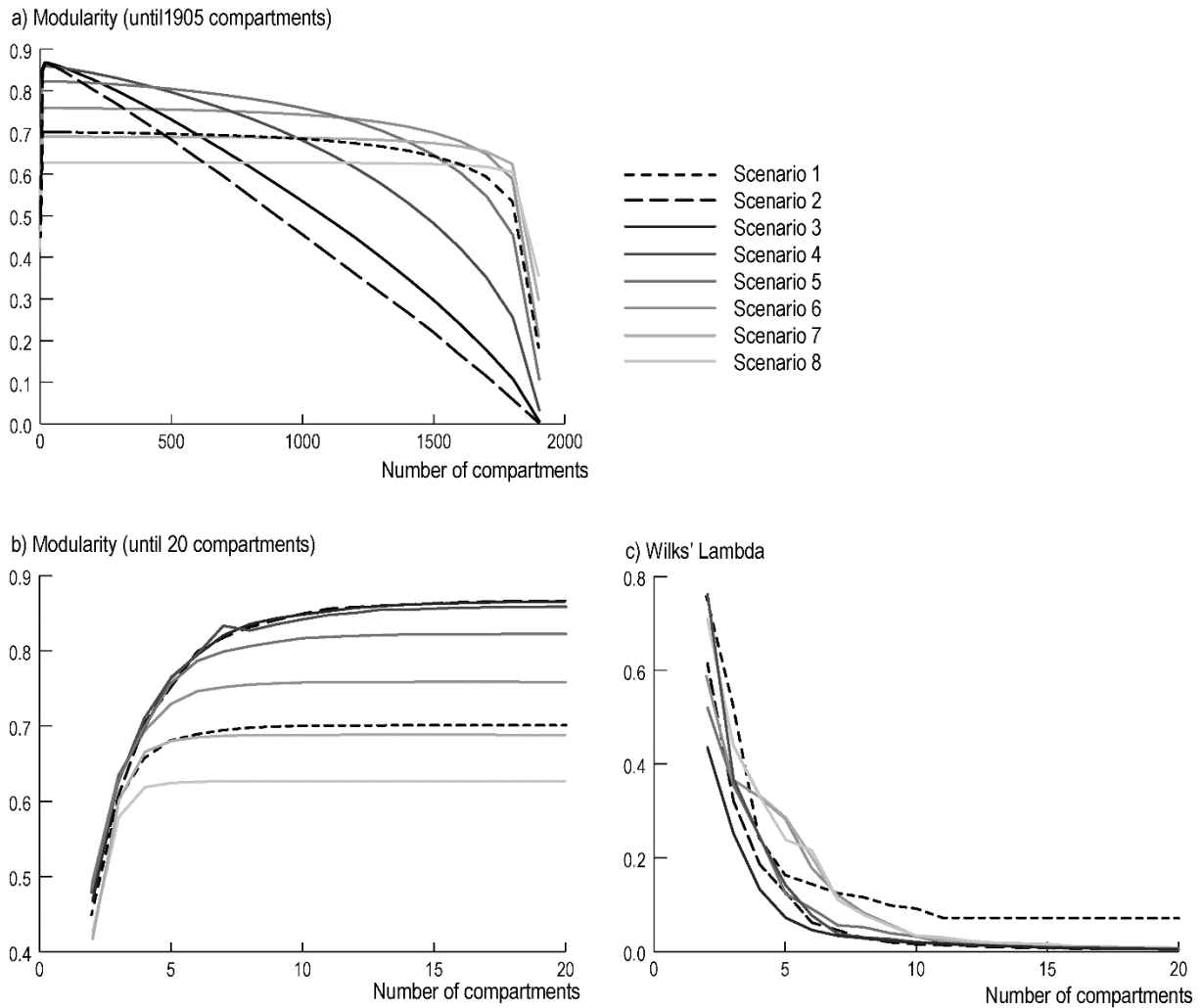


Figure 2 : Modularity (a, b) and Wilks' Lambda (c) according to the number of compartments for the eight scenarios of weighted adjacency matrix. Curves of modularity are drawn for all numbers of compartments in (a), and only until 20 compartments in (b), to make them comparable with Wilks' Lambda curves (c).

In a second step, the question was to determine whether modularity is a reliable indicator of the quality of the partitioning from a demographic point of view, with the aim of defining a suitable monitoring plan of water vole fluctuations. In this perspective, Wilks' Lambda values were calculated using 15 variables describing annual patch densities of water voles from 1989 to 2004. As the diffusion of water voles act as a travelling wave, strong collinearities occur between these variables. They were thus replaced by the orthogonal factors resulting from a principal component analysis. Using these factors, Wilks' Lambdas were computed for each weighting scenario and for each number of compartments from 2 to 20 (Fig. 2c). As expected, this criterion decreases as the number of compartments increases. Considering the 19 numbers of compartments between 2 and 20, Wilks' Lambda proves to be negatively correlated with modularity, Pearson's correlation coefficient being less than -0.85 ($p < 0.0001$) for all scenarios. Beyond this expected behavior, the relationship between Wilks' Lambda and modularity according to the weighting scenarios must be investigated for a fixed number of compartments. We therefore compared these criteria by drawing scatterplots for 5, 10, 15, and 20 compartments (Fig. 3).

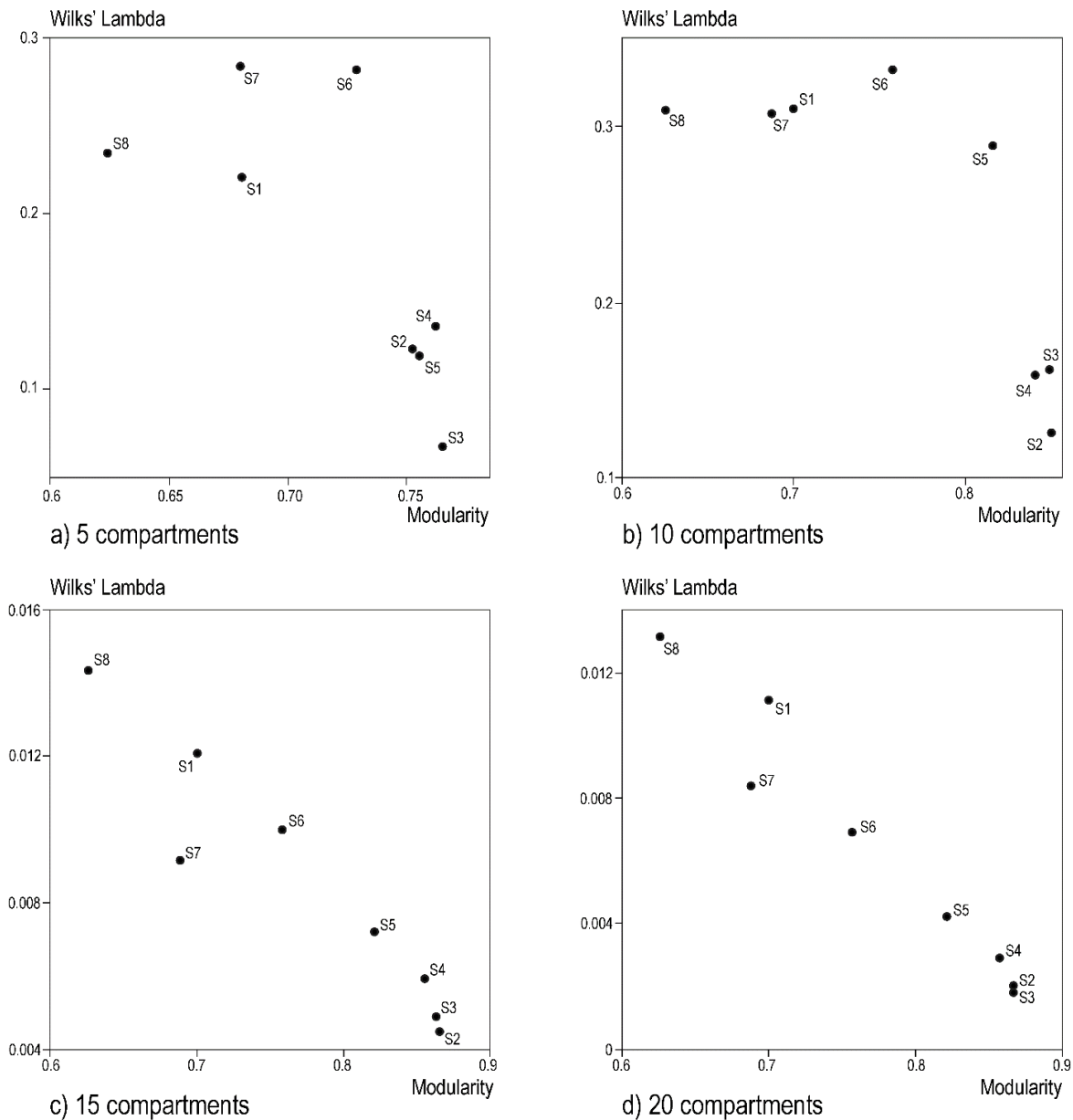


Figure 3 : Relation between modularity and Wilks' Lambda according to the number of compartments. For a small number of compartments (5 and 10), the relation proves to be negative but not linear, the group of scenarios with $\beta < 0$ being clearly more relevant. For 15 and 20 compartments, the relation remains negative but becomes linear, meaning that a larger modularity implies a partition more significant from a demographic point of view.

The scatterplots show a negative relationship between Wilks' Lambda and modularity. This relationship is not significant for five compartments, with a Pearson's correlation of -0.71 ($p = 0.052$). It becomes stronger and more significant as the number of compartments increases, with a correlation coefficient of -0.78 ($p = 0.022$) for 10 compartments, -0.95 ($p < 0.001$) for 15 compartments, and -0.98 ($p < 0.0001$) for 20 compartments.

In the following steps of the analysis, we focus on S2, which is considered to be one of the more relevant weighting scenarios given the previous results. The task is now to check the advantage of the compartmentalization to provide functionally homogeneous areas in comparison with (1) the compartments maximizing modularity using a graph thresholded at 1000 cost units and a non-weighted adjacency matrix (setting α and β to 0), and (2) the graph components resulting from a thresholding. For this last purpose, the MPG was used to generate graphs thresholded so as to contain

an increasing number of components from 2 to 24. We then compared the curves of Wilks' Lambda (Fig. 4a) and the g^2L criterion (Fig. 4b) according to the number of spatial units (compartments vs components) arising from these three methods of partition.

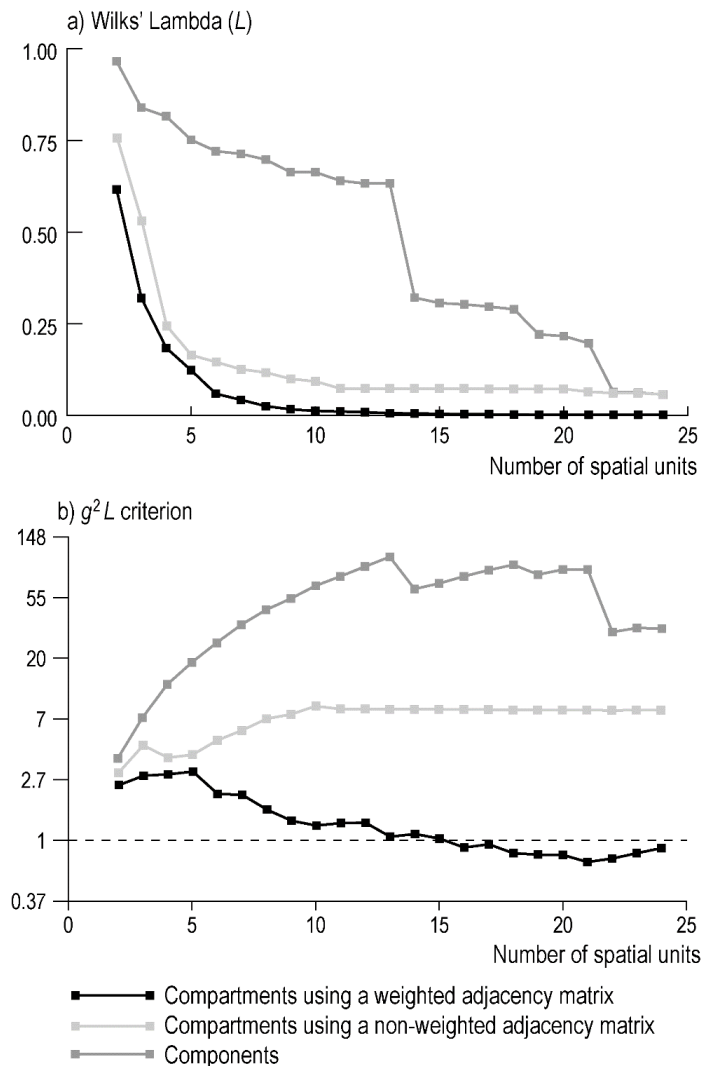


Figure 4 : Wilks' Lambda (a) and g^2L criterion (b) according to the number of areas for three methods of graph partitioning. A series of compartments was defined by maximizing modularity using the weighted adjacency matrix of scenario S2. Another series of compartments was defined from modularity without weighting. The components were defined by thresholding the links at different distances from the MPG. It should be noted that in case b, the y-axis is in a logarithm scale. For the g^2L criterion, the values below the dashed line (less than 1) correspond to the most relevant partitions (Marriott 1971).

Globally, Wilks' Lambda decreases while the number of compartments increases for all methods. However, the curve built from the components is largely above those built from the compartments. Moreover, the shape of the curves clearly differs, with a rapid decrease of Wilks' Lambda from 2 to 5 compartments for the clustering methods, contrasting with a plateau of high values up to 13 components for the thresholding method. The methods based on compartments exhibit curves of similar shapes, but for every number of compartments the Wilks' Lambda values are smaller with the weighted adjacency matrix. From 10 to 25 compartments both curves are stable, but the non-weighted adjacency matrix provides values always above 0.1 whereas the weighted method gives Wilks' Lambda lower than 0.005.

The difference between the three methods is more pronounced when investigating the curves of the g^2L criterion (Fig. 4b). The weighted compartmentalization is the only method that provides values decreasing as the number of compartments increases, until falling below 1. The values of g^2L computed from the non-weighted compartmentalization increase from 2 to 10 compartments then are stable around 7 when the number of compartments increases. Finally the values of g^2L provided by the component method strongly increase according to the number of components and almost exceed 10. A final step in the analysis is the choice of the number of compartments. The maximum value of modularity of 0.8672 is reached with 20 compartments, while the minimum value of the criterion g^2L (0.6878) is found with 21 compartments. Beyond these extreme values, the decrease of g^2L may also be used to find other relevant numbers of compartments, as shown in Fig. 5. The curve of this criterion plunges until a first level of 1.25 for 10 compartments, and then decreases again until the minimum value. The map of Fig. 5 illustrates these two nested partitions, set up without local optimization of modularity.

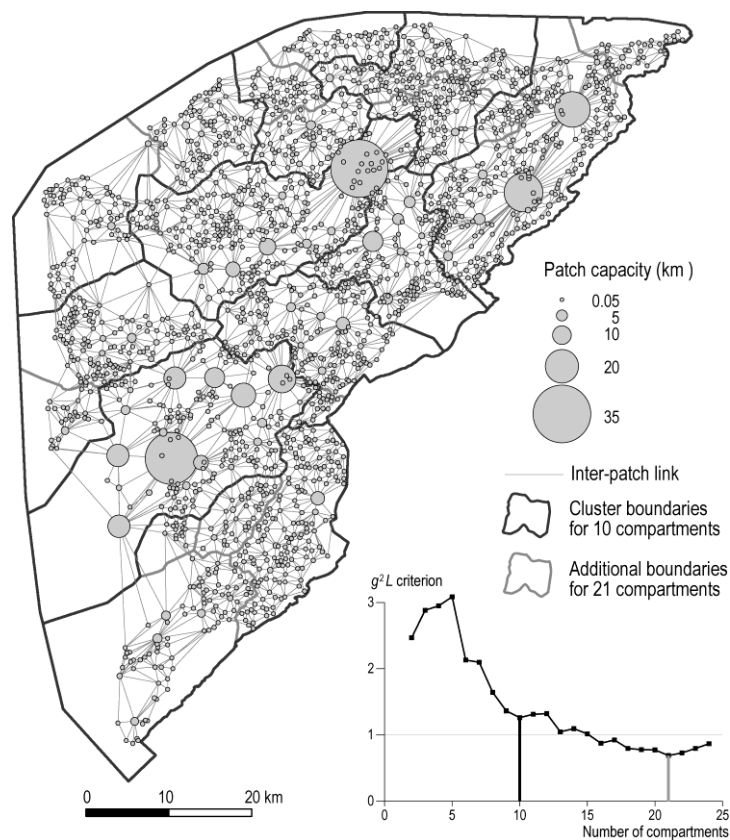


Figure 5 : Partition of the grassland network into 21 clusters. The number of clusters was selected from the maximum modularity value. It corresponds at the same time to the minimum value of the g^2L criterion. The cluster boundaries are determined by aggregating the Voronoi polygons of all the patches they include.

4. Discussion

Modularity is a criterion initially proposed for evaluating and generating partitions in a non-weighted graph. Its use is adapted to graphs in many domains, but in landscape ecology, it is restricted to specific types of landscape graphs. With the aim of defining a relevant partition into an ecological network, the method presented in this paper consists in adapting the computation of modularity to weighted graphs, including patch (node) capacity and weighting of the links in a weighted adjacency matrix. With this adaption, the binary links are replaced by continuous values expressing the strength of biological

fluxes between patches. Our proposition aims thus at providing a partition method that is well-adapted to graphs in which patches differ markedly in size (i.e. capacity) and links vary in distance. It is worth noting that landscape graphs do not always need to be weighted in that way. Our method does not apply to graphs with uniform nodes and binary links, which can be handled by the method of Gao et al. (2013). However, in other cases, we argue that the use of the weighted adjacency matrix is needed to take account of all the information contained in the basic elements of landscape graphs. The curves presented in Fig. 4 clearly show that in the present case study, the use of the weighted matrix provides more relevant compartments, and that whatever the weighting, compartmentalization methods are ecologically more relevant than components defined by link thresholding. These results thus confirm the value of the weighted adjacency matrix for graph compartmentalization.

A critical point is the setting of the weighted adjacency matrix, using the exponents α and β applied respectively to the distance characterizing the links and to the patch capacity. The first parameter is included in the function $p = e^{-\alpha d}$ transforming the distance d into probability p . This function is frequently met in models representing a process (e.g. dispersal) that decreases with distance, as in several patch-based connectivity metrics (Urban and Keitt 2001; Saura and Pascual-Hortal 2007; Foltête et al. 2012b). Usually, the parameter α is not directly set but defined so that a given probability corresponds to a key distance determined by users from their knowledge of the biology of the target species. In this study, we have chosen to apply this principle and then to vary only the parameter β by means of eight weighting scenarios. The results obtained showed that having exponent β equal to or greater than 1 (scenarios 6, 7, and 8) gives weaker modularity. Setting β to 0 (scenario 1) also leads to quite low modularity values. Conversely, having exponent β less than 1 but not equal to 0 (scenarios 2, 3, 4 and 5) provides high modularity values. This outcome confirms that patch capacity should be taken into account to properly quantify the interactions between patches, but with a small influence relative to inter-patch distance. It should, however, be noted that the best result obtained with scenario 2 ($\beta = 0.1$) has no overall value and may depend on the case study. Scenario 4 ($\beta = 0.5$) provides results that are marginally lower than in scenarios 2 and 3 but could be used by default, because this scenario involves a square root applied to the product of two areas, that is, patch capacities are simply weighted by the unit area.

From the data describing the demographic fluctuations of the montane water vole, statistical analyses based on Wilks' Lambda were set up to investigate the relevance of the compartmentalization under eight weighting scenarios. The results show a strong (and negative) relationship between modularity and Wilks' Lambda, meaning that for a given number of compartments, a higher modularity value is found for compartments that are more homogeneous in terms of demographic fluctuations. Likewise, the optimal number of compartments was almost the same given the maximum value of modularity (20 compartments) and the minimum value of g^2L (21 compartments). These outcomes confirm the relevance of modularity in an empirical framework, thus supplementing the result obtained by Gao et al. (2013) for theoretical properties of the expected topology of compartments considered relevant. Such a result is interesting when modularity will be the only indicator of compartmentalization quality (e.g., where demographic or genetic data are unavailable or too sparse to make a statistical analysis). In practical terms, for a user seeking to delineate compartments but having no idea about the best way to define the weighted adjacency matrix (and especially the parameter β), modularity can therefore be used as a guide for identifying a suitable setting after a step of test and trials.

Another issue to address is the ability of the weighted adjacency matrix to properly quantify the potential fluxes between patches. The earliest works on landscape graphs already took the notion of flux into consideration (Urban and Keitt 2001) and mentioned its asymmetric character. Other studies have also outlined this asymmetry of dispersal flux between patches (Gustafson and Gardner 1996) and the limit of the approaches focused on the distance effect alone (Bodea et al. 2008). Consequently, whenever habitat patches differ in weight (i.e. their demographic potential), dispersal flux should be considered an asymmetric process. However, modularity relies on inter-patch connections alone and cannot include the direction of potential fluxes. It could be argued that in this area, the method based on the weighted adjacency matrix is consistent with most patch-based graphs which are designed with undirected links (Galpern et al. 2011). Moreover, this matrix expresses the spatial interactions

between patches in a form similar to the probability of connectivity index (Saura and Pascual-Hortal 2007), a widely used connectivity metric. But in any case, since only the notion of connection matters in the compartmentalization approach, this method does not suffer from the simplification of fluxes by undirected weighted links.

As compartments defined from a landscape graph are expected to be areas with more internal than external exchanges, one perspective of this work could be to investigate landscape genetics issues. Compartmentalization should be clearly related to genetic clustering and the simplest way to test such a connection would be to compare delineations from both approaches. Moreover, since the modularity-based compartmentalization results from an agglomerative process generating nested partitions, the method could be used for exploring the effect of landscape on gene fluxes at several spatial scales, following the initial work of Galpern et al. (2012). In their study, those authors seek how to investigate the effect of isolation by resistance on the target species. They generate multiple “grains of connectivity” by starting from the most precise landscape map and progressively agglomerating the graph nodes via the Voronoi tessellation underlying the minimal planar graph. Their statistical results show that the finest spatial grain is not the best way to highlight the effect of landscape on genetic structure. Since the agglomerative process was based on the distance criterion alone, the nested partitions generated by maximizing modularity could be an interesting alternative for exploring spatial scales while including potential fluxes (and not just spatial distance) in the analysis.

5. Conclusion

The adaptation proposed for computing modularity proves to be suitable for defining compartments from weighted landscape graphs. The inclusion of a weighted adjacency matrix allows the user to take account of patch capacity (mostly patch area) and inter-patch distance, whatever the type of distance characterizing the graph links. The setting of this matrix by means of two parameters depends on each case study, but given the significant relation found between modularity and Wilks’ Lambda (i.e. quality of the statistical partition for multivariate data), modularity can be used as a reference criterion to explore several possibilities and approach the optimal partition.

Acknowledgements

This research has been funded by the French Ministry of Ecology (Campagraphe project in DIVA3 Program). Computations were performed on the supercomputer facilities of the “Mésocentre de calcul de Franche-Comté”. This research is part of the Jurassian Arc long-term ecological research site (<http://zaaj.univ-fcomte.fr/?lang=en>).

References

- Awade M, Boscolo D, Metzger JP (2012) Using binary and probabilistic habitat availability indices derived from graph theory to model bird occurrence in fragmented forests. *Landscape Ecol* 27:185–198
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci USA* 101:3747–3752
- Berthier K, Piry S, Cosson JF, Giraudoux P, Foltête JC, Defaut R, Truchetet D, Lambin X (2014) Dispersal, landscape and travelling waves in cyclic vole populations. *Ecol Lett* 17:53–64
- Berthier K, Chaval Y, Galan M, Charbonnel N, Cosson JF (2009) Dispersion individuelle : Conséquences, in: Delattre P, Giraudoux P (Eds.), *Le campagnol terrestre : prévention et contrôle des populations*. QUAE Edition, Versailles, pp. 39–47
- Bjornstad ON, Ims RA, Lambin X (1999) Spatial population dynamics: analyzing patterns and processes of population synchrony. *Trends Ecol Evol* 14:427–432
- Blant MA, Beuret B, Poitry R, Joseph E (2009) Influence of landscape and soil on the intensity of pullulations of vole (*Arvicola terrestris scherman*) in Swiss Jura. *Rev Suisse Agric* 41:301–307
- Bodea M, Burrage K, Possingham HP (2008) Using complex network metrics to predict the persistence of metapopulations with asymmetric connectivity patterns. *Ecol Model* 214:201–209
- Bodin O, Norberg J (2007) A network approach for analyzing spatially structured populations in fragmented landscape. *Landscape Ecol* 22:31–44
- Brandes U, Delling D, Gaertler M, Görke R, Hoefer M, Nikoloski Z, Wagner D (2008) On modularity clustering. *IEEE Trans Knowl Data Eng* 20:172–188
- Cavanaugh KC, Siegel DA, Raimondi PT, Alberto F (2014) Patch definition in metapopulation analysis: a graph theory approach to solve the mega-patch problem. *Ecology* 95:316–328
- Clauzel C, Bannwarth C, Foltête JC (2015) A planning tool for integrating broad-scale connectivity in habitat restoration: an application to pond creation in eastern France. *J Nat Conserv* 23:98–107
- Duhamel R, Quéré JP, Delattre P, Giraudoux P (2000) Landscape effects on the population dynamics of the fossorial form of the water vole (*Arvicola terrestris sherman*). *Landscape Ecol* 15:89–98
- Everitt BS, Dunn G (1991) *Applied Multivariate Data Analysis*. Edward Arnold, London.
- Fall A, Fortin MJ, Manseau M, O'Brien D (2007) Spatial graphs: principles and applications for habitat connectivity. *Ecosystems* 10:448–461
- Foltête JC, Clauzel C, Vuidel G, Tournant P (2012a) Integrating graph-based connectivity metrics into species distribution models. *Landscape Ecol* 27:557–569
- Foltête JC, Giraudoux P (2012) A graph-based approach to investigating the influence of the landscape on population spread processes. *Ecol Ind* 18:684–692
- Foltête JC, Vuidel G, Clauzel C (2012b) A software tool dedicated to the modelling of landscape networks. *Env Model Soft* 38:316–327
- Foltête JC, Girardet X, Clauzel C (2014) A methodological framework for the use of landscape graphs in land-use planning. *Landscape Urban Plan* 124:140–150
- Foltête JC, Couval G, Fontanier M, Vuidel G, Giraudoux P (2016) A graph-based approach to defend agro-ecological systems against water vole outbreaks. *Ecol Ind* 71:87–98
- Gao P, Kupfer JA, Guo D, Lei TL (2013) Identifying functionally connected habitat compartments with a novel regionalization technique. *Landscape Ecol* 28:1949–1959
- Galpern P, Manseau M, Fall A (2011) Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biol Conserv* 144:44–55

- Galpern P, Manseau M, Wilson P (2012) Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Mol Ecol* 21:3996–4009
- Giraudoux P, Delattre P, Habert M, Quéré JP, Deblay S, Defaut R, Duhamel R, Moissenet MF, Salvi D, Truchetet D (1997) Population dynamics of fossorial water vole: a land use and landscape perspective. *Agric Ecosyst Environ* 66:47–60
- Giraudoux P, Pradier B, Delattre P, Deblay S, Salvi D, Defaut R (1995) Estimation of water vole abundance by using surface indices. *Acta Theriol* 40:77–96
- Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99:7821–7826
- Gustafson EJ, Gardner RH (1996) The effect of landscape heterogeneity on the probability of patch colonization. *Ecology* 77:94–107
- Hall AL, Beissinger SR (2014) A practical toolbox for design and analysis of landscape genetics studies. *Landscape Ecol* 29:1487–1504
- Luque S, Saura S, Fortin MJ (2012) Landscape connectivity analysis for conservation: insights from combining new methods with ecological and genetic data. *Landscape Ecol* 27:153–157
- Lookingbill TR, Elmore AJ, Engelhardt KAM, Churchill JB, Gates E, Johnson JB (2010) Influence of wetland networks on bat activity in mixed-use landscapes. *Biol Conserv* 143:974–983
- McLachlan GJ (2004) *Discriminant Analysis and Statistical Pattern Recognition*. Wiley
- Marriott FHC (1971) Practical problems in a method of cluster analysis. *Biometrics* 27:501–514
- Minor ES, Urban DL (2007) Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecol Appl* 17:1771–1782
- Moilanen A (2011) On the limitations of graph-theoretic connectivity in spatial ecology and conservation. *J Appl Ecol* 48:1543–1547
- Morilhat C, Bernard N, Foltête JC, Giraudoux P (2008) Neighbourhood landscape effect on population kinetics of the fossorial water vole (*Arvicola terrestris scherman*). *Landscape Ecol* 23:569–579
- Newman M, Girvan M (2004a) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113
- Newman M (2004b) Analysis of weighted networks. *Phys Rev E* 70:056131
- Saura S, Pascual-Hortal L (2007) A new habitat availability index to integrate connectivity in landscape conservation planning: Comparison with existing indices and application to a case study. *Landscape Urban Plann* 83:91–103
- Schuetz P, Caflich A (2008) Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement. *Phys Rev E* 77:046,112
- Taylor P, Fahrig L, With W (2006) Landscape connectivity: A return to basics. In: Crooks KR, Sanjayan M (ed), *Connectivity Conservation*. Cambridge University Press, Cambridge, pp29–43
- Tournant P, Afonso E, Giraudoux P, Roué S, Foltête JC (2013) Evaluating the effect of habitat connectivity on the distribution of lesser horseshoe bat maternity roosts using landscape graphs. *Biol Conserv* 164:39–49
- Urban DL, Keitt TH (2001) Landscape connectivity: a graph theoretic approach. *Ecology* 82:1205–1218
- Urban DL, Minor ES, Treml EA, Schick RS (2009) Graph models of land mosaics. *Ecol Lett* 12:260–273
- Vogt P, Riitters KH, Iwanowski M, Estreguil C, Kozak J, Wade TG, Wickham JD (2007) Mapping spatial patterns with morphological image processing. *Landscape Ecol* 22:171–177
- Webster R (1972) Wilks's criterion: a measure for comparing the value of general purpose soil classifications. *J Soil Science* 11:254–260