

Multi-dimensional Urban Network Percolation

Juste Raimbault

▶ **To cite this version:**

Juste Raimbault. Multi-dimensional Urban Network Percolation. Journal of Interdisciplinary Methodologies and Issues in Science, 2019, Analysis of networks and graphs, 10.18713/JIMIS-180719-5-5 . halshs-02361292

HAL Id: halshs-02361292
https://shs.hal.science/halshs-02361292

Submitted on 13 Nov 2019

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.

JOURNAL OF INTERDISCIPLINARY METHODOLOGIES AND ISSUES IN SCIENCE

Multi-dimensional Urban Network Percolation

Juste RAIMBAULT[1,2,3,*]

[1]UPS CNRS 3611 ISC-PIF, France
[2]CASA, UCL, UK
[3]UMR CNRS 8504 Géographie-cités

*Corresponding author: juste.raimbault@polytechnique.edu

DOI: 10.18713/JIMIS-180719-5-5
Submitted: March 17, 2019 - Published: October 23, 2019
Volume: **5** - Year: **2019**
Issue: **Graph and network analysis**
Editor: *Vincent Labatut*

Abstract

Network percolation has recently been proposed as a method to characterize the hierarchical structure of an urban system from the bottom-up. This paper proposes to extend urban network percolation in a multi-dimensional way, to take into account both urban form (spatial distribution of population) and urban functions (here as properties of transportation networks). The method is applied to the European urban system to reconstruct endogenous urban regions. The variable parametrization allows to consider patterns of optimization for two stylized contradictory sustainability indicators (economic performance and greenhouse gases emissions). This suggests a customizable spatial design of policies to develop sustainable territories.

Keywords

Road network; Multi-dimensional percolation; European urban system; Mega-city region

I INTRODUCTION

1.1 Towards multidimensional urban network percolation

The structure of road networks can be used as a proxy to understand its past growth dynamics, but also has a significant impact on the future sustainability of territories it irrigates. Diverse methods to characterize the structure of spatial networks, and more particularly road networks, have been developed in that context. They include classical network indicators such as centralities (Crucitti *et al.*, 2006) but also more elaborated constructions capturing more realistic processes in terms of street network use (Lagesse *et al.*, 2015). Such studies of urban networks are by essence interdisciplinary, or at least imply complementary viewpoints from diverse disciplines. These for example include architecture with space syntax (Hillier *et al.*, 1976), physics

©JIMIS, Creative Commons

with the study of spatial networks (Barthélemy, 2011), or social science disciplines concerned with space such as geography (Ducruet and Beauguitte, 2014).

A method to characterize the hierarchical structure of such urban spatial networks is network percolation, initially applied to urban road networks by Arcaute *et al.* (2016). Percolation in physics can be understood in a broad sense as processes related to the progressive occupation or connection of nodes of a network. It is generally associated to a phase transition with the emergence of a giant cluster at a given probability of connection (Stauffer and Aharony, 2014). Practical applications in different fields include the quantification of network robustness (Callaway *et al.*, 2000) or the modeling of epidemic spreading (Newman and Watts, 1999).

Such approaches have been applied to urban systems with other applications than the study of networks. Makse *et al.* (1998) model urban growth with a local percolation model for site occupancy. Arcaute *et al.* (2016) focus on the analysis of street networks and extract endogenous urban regions for UK which correlate with socio-economic properties, and provide a definition of urban areas which highly correlates with land-cover data. Piovani *et al.* (2017) apply road network percolation at the mesoscopic scale of London metropolitan area, in relation with a retail location model. At a larger scale, the paradigm of percolation transition has been applied to the study of urban traffic dynamics (Li *et al.*, 2015; Zeng *et al.*, 2019). In spatial statistics, this method can be used to characterize the spatial morphology of point patterns (Huynh *et al.*, 2018).

Existing heuristics generally focus on a single dimension or property of the urban system. However, such systems are known to be highly multidimensional. For example, the morphological dimension of networks is complementary to the functional properties of the urban environment (Burger and Meijers, 2012). The link between urban form and function remains in particular an open question (Batty and Longley, 1994). More generally, the inclusion of multiple dimensions in urban analysis is still a research direction to be investigated, as in the case of agent-based models for example (Perez *et al.*, 2016). This paper addresses such a gap in the case of urban network percolation, by introducing a multi-dimensional percolation heuristic. The method allows combining different dimensions of the urban system, the same way that Cottineau *et al.* (2018) link population density with commuting flows to produce multiple definitions of urban areas. We can indeed expect significantly different qualitative behaviors when switching from a single-dimension percolation to a multi-dimensional one, since the spatial structure of different urban dimensions are correlated but also largely complementary. If this is the case, the method then captures more aspects of the urban system and some of their interrelations.

1.2 Sustainability of mega-urban regions

Beside these methodological issues of characterizing urban networks and more particularly their endogenous hierarchical structure, some related applied research issues can be considered. Indeed, quantitative tools are needed to evaluate the sustainability of recently emerged urban forms. In particular, according to Le Néchet (2017), the most recent transition of human settlement systems (in the sense of Sanders (2017), i.e. a change in the dynamical regime ruling the evolution of the spatial structure of settlements) is the emergence of mega-city regions. These have been defined by Hall and Pain (2006) as polycentric urban structures highly integrated in terms of flows. The transition imply complex processes such as changes in the governance structure, and can not be associated to the stylized transition identified by Louf and Barthelemy (2013) in a toy urban model including negative externalities of congestion only.

To what extent these new urban forms are sustainable, for example in the broad sense of UN development goals (Komiyama and Takeuchi, 2006), remains an open question. Indeed, these integrated mega-city regions may for example imply different patterns of economic and transportation flows and thus exhibit various performances regarding different indicators of sustainability.

Case studies of targeted megacities have for example focused on the links between urban form and mobility or resources management (Sorensen and Okata, 2010). A significant literature has focused on the sustainability of mega urban regions with a qualitative approach, such as Laquian (2011) which establishes a typology of these and corresponding governance structures in the Asian context. In the case of Europe, Marull *et al.* (2013) use econometric analysis to see the economic and ecological advantage of integrated urban regions. Feng *et al.* (2018) introduce a method to measure the level of polycentricity of an urban mega-region. Su *et al.* (2017) propose a framework to evaluate the performance of urban mega-regions, regarding economic, environmental, social and spatial dimensions.

Beside these targeted studies, a remaining open issue, to the best of our knowledge never tackled at this small geographical scale, is the endogenous characterization of such urban regions. The delineation of these geographical systems is often taken as exogenous, and their performance and sustainability is then evaluated. Several approaches can be taken to define such systems, including integration through transportation networks, continuity of night lights, and economic productivity thresholds (Lang and Dhavale, 2005; Florida *et al.*, 2008). In these existing studies, parameters to define such regions remain fixed, and an endogenous hierarchical structure is possibly ignored. The second component of our research question will thus be the endogenous characterization of urban regions and how this can be applied to study their sustainability.

1.3 Proposed approach

This work proposes to partly tackle the two above research questions by linking them. More precisely, it first investigates from a methodological viewpoint how urban network percolation can be generalized to multiple dimensions, and secondly explores the endogenous characterization of mega-city regions using such a multi-dimensional percolation, and how this can be used to quantify the sustainability of these systems. These two axis are tightly linked. Indeed, on the one hand street network percolation has initially been proposed to identify endogenous entities in territorial systems, and on the other hand mega-urban regions are characterized simultaneously by morphological dimensions (continuity of the built environment) and functional dimensions (high integration of flows).

We therefore consider percolation on two dimensions of the urban systems, one linked to urban form which is the spatial distribution of population, and one linked to transportation networks, through structural indicators quantifying local road networks. We apply in particular the heuristic to urban morphology and road network topology measures in Europe. The idea to combine urban form with network topology measures relies on the capture of the link between urban form and function as already mentioned, urban functions being assumed as distributed by transportation networks (Raimbault, 2018b). In general, the interactions between transportation networks and territories have been shown to be central processes in urban dynamics (Offner *et al.*, 2014).

We thus introduce in this paper a multi-dimensional percolation method, and apply it to the endogenous characterization of urban regions, to finally compute stylized sustainability indicators on the constructed regions. These sustainability measures are extrapolated from gravity flows, and are used as a proof-of-concept of how this work can be applied towards policy-making. As

detailed below, we use paradigmatic opposed dimensions of sustainability which are greenhouse gases emissions and economic productivity.

The originality of our contribution relies thus on several points: (i) this is to the best of our knowledge the first time a multi-dimensional percolation method is applied to urban systems; (ii) we furthermore apply it on the significant spatial extent of all European Union; and (iii) we link the clusters obtained with simple sustainability measures.

The rest of this paper is organized as follows: we first describe the multi-dimensional percolation heuristic, the data and variables to which it is applied and how the network is constructed, and the indicators used to characterize the sustainability of clusters produced. We then describe the results of applying this method to population and road network indicators for the whole European Union, focusing on the endogenous regions produced and their sustainability properties. We finally discuss possible developments and the implications of this methodology to the design of policies.

II METHODS

2.1 Multi-dimensional percolation

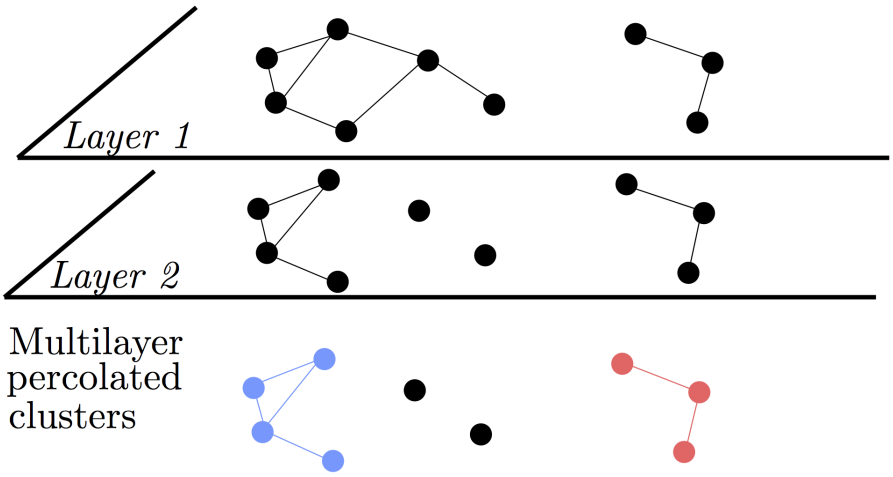


Figure 1: Schematic representation of the multi-dimensional network percolation heuristic. Two layers are considered here, with similar nodes but different links in each. To each layer is associated a node variable. In each layer, links are created according to a distance threshold (percolation radius) and a variable value threshold. The final clusters (bottom in color) are obtained by considering the links present in each layer.

2.1.1 Proposed heuristic

Percolation processes in multilayer networks have been proposed as an extension within simple networks (Boccaletti *et al.*, 2014). A generalization of epidemic spreading can for example be achieved using this framework (Son *et al.*, 2012). In the case of multilayer networks sharing the same nodes for all layers, often called multiplex networks (Aleta and Y., 2018), bond percolation has also been studied (Hackett *et al.*, 2016). In the case of our heuristic, bond percolation is operated between two cells given a distance threshold, and furthermore with a threshold parameter for each layer assuming a node function within each layer. The distance-based connection is similar to generative processes for random euclidian networks (Penrose, 1999).

We do not call our method "multi-layer percolation", since nodes are common between layers. We show in Figure 1 for a schematic representation of the method. It can be implemented with

a propagation heuristic or directly working on adjacency matrices. The rationale behind combining a thresholding for each layer variable with a distance thresholding relies on the idea that a first component for two points to interact is a geographical proximity. The second component is a strong enough intensity of the urban dimension captured by each layer, simultaneously for all layers considered. This recalls Tobler's first law of geography (Tobler, 2004) in a multi-dimensional way.

2.1.2 Formal description

More formally, let assume a set of nodes $V = v_i$ common to all layers of the network, and layers edges E_j for layer j taken as empty at the initial state of the algorithm. Each node has a value of the considered variables associated to each layer, written as a matrix v_{ij}. The algorithm works as follows:

1. Percolation is first done within each layer. For each layer, a link $e_{kl} \in E_j$ is created if $d(v_k, v_l) < r_0$ where d is the distance between the nodes (which can be any distance) and r_0 the percolation radius, and if $v_{kj} > \theta_j$ and $v_{lj} > \theta_j$ where θ_j is the threshold for layer j.
2. Layers are combined, by computing the final percolated network edges E as the links contained within all layers simultaneously. The multi-dimensional percolation clusters are then the connected components of this network (V, E).

The parameters implied in this heuristic are the percolation radius r_0 and the variable thresholds $\vec{\theta} = \theta_j$ for each layer j. Varying these parameters allows considering different levels of percolation.

2.1.3 Application to gridded networks

The previous generic method must be applied to a consistent urban network, in the sense of multiple variables associated to spatial nodes. To each variable then corresponds a network layer. We propose to work on gridded networks, namely nodes regularly spaced in two dimensions. In practice, these nodes will be the center of raster cells. We will consider two layers, one defined by population density and the other by road network structure indicators computed at the same resolution.

2.2 Empirical data and network construction

We detail now how the empirical layers were computed and the network constructed. A grid of population density morphology indicators has been computed on spatial moving windows of width 50km for all European Union by Raimbault (2018a), with an offset resolution of 5km. From this study we get the aggregated population, producing a raster grid of population with a resolution of 5km.

Furthermore, road network topology indicators were computed at a similar resolution by Raimbault (2019). In practice, (i) the full OpenStreetMap road network for Europe was simplified at the minimal resolution of 100m, keeping the topological properties and link attributes (including maximal speed for example); (ii) for each cell of the population morphology raster, the road network within the same spatial window of width 50km was retrieved; (iii) network structure indicators (summary statistics, centralities, etc.) were computed on this local network, providing a network indicators raster at the same resolution than population.

We use this data to construct a two layers abstract network: a layer which variable is given by population density, and a second layer which variable is given by a road network indicator.

Nodes are the center of cells (thus disposed in space on a grid of step 5km). We test several possible networks by varying the road indicator taken into account for the second layer. In particular, we test the following indicators:

1. number of edges N_E;
2. number of vertices N_V;
3. cyclomatic number μ which is defined by $\mu = N_E - N_V + c$ where c is the number of connected components of the graph; this measure captures the number of distinct cycles in the graph;
4. euclidian efficiency v_0, defined by Banos and Genre-Grandpierre (2012), as the average rate between network distance and euclidian distance for all couples of links.

The choice of these measures is aimed at capturing basic aspects of network structure, and functional properties especially for the two last. Indeed, euclidian efficiency measures how the network is performant to link nodes, while the number of cycles is linked to redundancy of paths and in a way to robustness. These choices are arbitrary, but several aspects of transportation networks are still captured by these indicators. An increase of each is related to a more performant network regarding different dimensions, what is relevant for our application. A systematic study with other indicators such as centralities or accessibilities is out of the scope of this paper.

The percolation on such an abstract network is a necessary condition in our case to link the different dimensions considered, namely population distribution and local road network properties. We have therefore two levels of networks in our approach, namely the physical road network which local properties are taken here as input, and the abstract two layer network on which we do the percolation.

We will in the following write θ_P for the threshold parameter of the population layer, and θ_N for the threshold parameter of the network layer. In practice, these parameters will be given in the following as quantile level of the corresponding variable, for an easier interpretation and conception of experience plans. The name of the road network indicator considered will be written v_N.

2.3 Sustainability indicators

As already detailed, urban regions may perform more or less well regarding different dimensions of sustainability. We propose to use the endogenous definition of regional urban systems produced by the percolation algorithm to evaluate their sustainability, in terms of conflicting objectives of economic integration and greenhouse gases emissions. The definition of sustainability, or sustainable development, is by essence multi-dimensional (Viguié and Hallegatte, 2012). Its characterization as quantitative indicators is even more subject to numerous degrees of freedom. We work here with these two stylized indicators for two conflicting dimensions, as a proof-of-concept. These dimensions can furthermore be approximated indirectly from gravity models as we will describe below. By introducing other datasets, our work could be extended to more realistic indicators and other dimensions.

We use the EDGAR database (Janssens-Maenhout *et al.*, 2017) (version 4.3.2) for local grid estimates of greenhouse gases emissions. We use the latest year available, namely 2012. As its resolution is much smaller than our indicator grid, we aggregate the emissions on the closer indicator point for each cell of the emission database. Since according to Lashof and Ahuja (1990) most of the greenhouse effect is caused by CO_2, and as in terms of emissions in the database we find that it represents 98.2% in mass proportion of all gases, we only consider it.

Applying a gravity model to each region, we estimate abstract transportation flows within each with a gravity model. More precisely, following Raimbault (2018c), a potential flow between two points i and j can be estimated with the following expression

$$\phi_{ij}^{(k)} = \left(\frac{v_i^{(k)} v_j^{(k)}}{(\sum_l v_l)^2} \right)^{\gamma} \cdot \exp\left(\frac{-d_{ij}}{d_0} \right) \tag{1}$$

where $v_i^{(k)}$ are either population or effective local GHG emissions computed with the EDGAR database (indexed by $k = 1, 2$ respectively), d_{ij} the distance between the two points, d_0 a distance decay parameter, and γ a scaling parameter. Indeed, the economic activity follows relatively well scaling laws of populations (Bettencourt *et al.*, 2007), the exponent being dependant on the activity and the definition of areas on which it is estimated (Cottineau *et al.*, 2017). The distance decay captures the geographical span of potential interactions. These two parameters γ, d_0 are left free and varying them allows considering stylized configurations, such as long or short span interactions, or infra- or supra-linear scaling activities. Finally, considering the flow with the population variable ($k = 1$) provides a proxy for economic flows, while with the GHG emissions ($k = 2$) it provides a proxy for emissions in relation with this economic activity. Indeed, effective emissions are linked to local emissions and transportation emissions linked to the intensity of flows.

We then consider the sum of all these flows of points within the geographical span of a given cluster of nodes in our network. These clusters are obtained with the percolation method described above, and are numbered by $1 \leq c \leq C$. For the sake of simplicity, we approximate the corresponding geographical area as the convex Hull envelope of the points in the cluster, that we write K_c. By cumulating the flows, we therefore define the total economic flow as the sum of economic flows by

$$E_c = \sum_{i,j \in K_c} \phi_{ij}^{(1)} \tag{2}$$

and the total emissions due to flows by

$$G_c = \sum_{i,j \in C_c} \phi_{ij}^{(2)} \tag{3}$$

These indicators have no interpretable unit and need to be renormalized. This allows defining a relative economic inefficiency by

$$e_c = 1 - \frac{\max_c E_c - E_c}{\max_c E_c - \min_c E_c} \tag{4}$$

where $\max_c E_c$ (resp. $\min_c E_c$) is the maximal (resp. minimal) value of E_c across all clusters. We define in a similar way the relative potential emissions by

$$g_c = \frac{\max_c G_c - G_c}{\max_c G_c - \min_c G_c} \tag{5}$$

where $\max_c G_c$ (resp. $\min_c G_c$) is the maximal (resp. minimal) value of G_c across all clusters.

Both indicators should be minimized for sustainability. Their value is dependant on the number of clusters and their extent, i.e. the geographical surface they cover. To be able to compare values across different clusterings (corresponding to different parameter values for the percolation heuristic), we finally define normalized indicators $\tilde{e}_c, \tilde{g}_c$ in a similar way, but the extrema being computed on all other possible urban configurations with the same γ, d_0 values.

Using these potential flows follows the logic of Arbabi *et al.* (2019) which show a need for improved intra-city-region mobility in England and Wales. Considering the regions as entities in which such transportation development policies can more easily been developed, we look at the sustainability of different possible regions if these potential flows were realized. Varying the parameters γ and d_0 allows controlling for the economic activity considered (high γ values correspond to high added-value activities) and the span of interactions through d_0.

For descriptive purposes, we also consider summary measures of clusters, as the population P_c and effective emissions EM_c taken as the sum of population (resp. emissions) of the points in K_c.

III RESULTS

3.1 Implementation

In practice, the analysis is implemented using R and the igraph package. Source code, data and results are available on the open git repository of the project at `https://github.com/JusteRaimbault/UrbanMorphology`. The network is constructed by superposing the population density layer with the network layer, starting from the 5km resolution spatial fields for morphological and network indicators. This network is filtered with the threshold parameters for each layer and with the radius parameter. Connected components yield the clusters that we interpret as endogenous regions.

3.2 Percolation transition and fractal dimension

In its application to road networks by Arcaute *et al.* (2016), the structure of the national urban system for UK is captured by studying the percolation transition, i.e. the variation of the size of the largest cluster as a function of the percolation radius. As this signature is tightly linked to historical, cultural and geographical conditions, the application to different urban systems should yield different results. We study here this property, for different threshold parameter values. We make the radius vary between 8km and 100km with a one km step, have a fixed population threshold $\theta_P = 0.85$, test all road network indicators, and three network layer thresholds $\theta_N \in \{0; 0.8; 0.95\}$. These values yield a good precision for the radius which is the most important variable to study transition and estimate fractal dimensions, while this population threshold is enough to provide a broad coverage for large radiuses (as shown by other explorations described below). Changing network variables and their threshold allows us to investigate how the transition behavior does change regarding the dimension considered and its intensity.

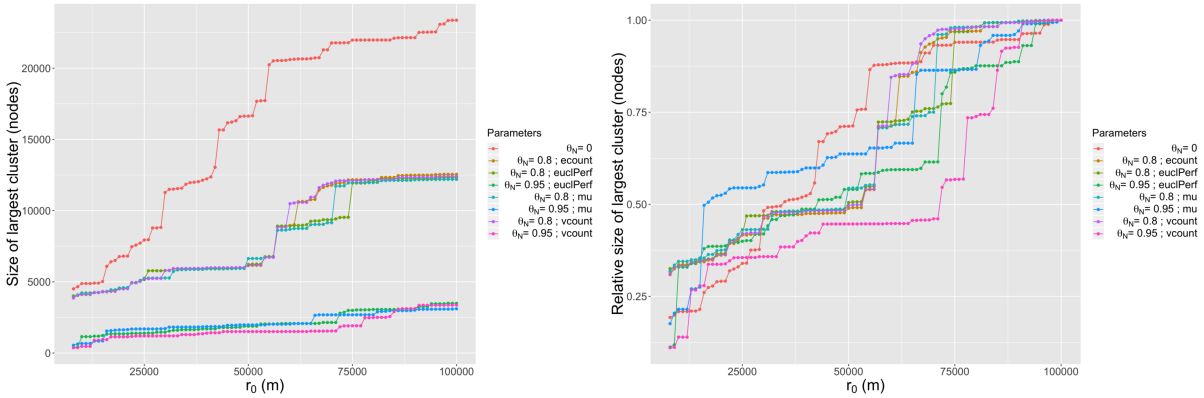


Figure 2: Percolation transition. On the left, we plot the size of the largest cluster in each configuration in terms of nodes, as a function of the percolation radius r_0. Color gives the other percolation parameters. On the right, the plot is similar but with the size relative to the size of the largest cluster obtained with the maximal radius in each configuration.

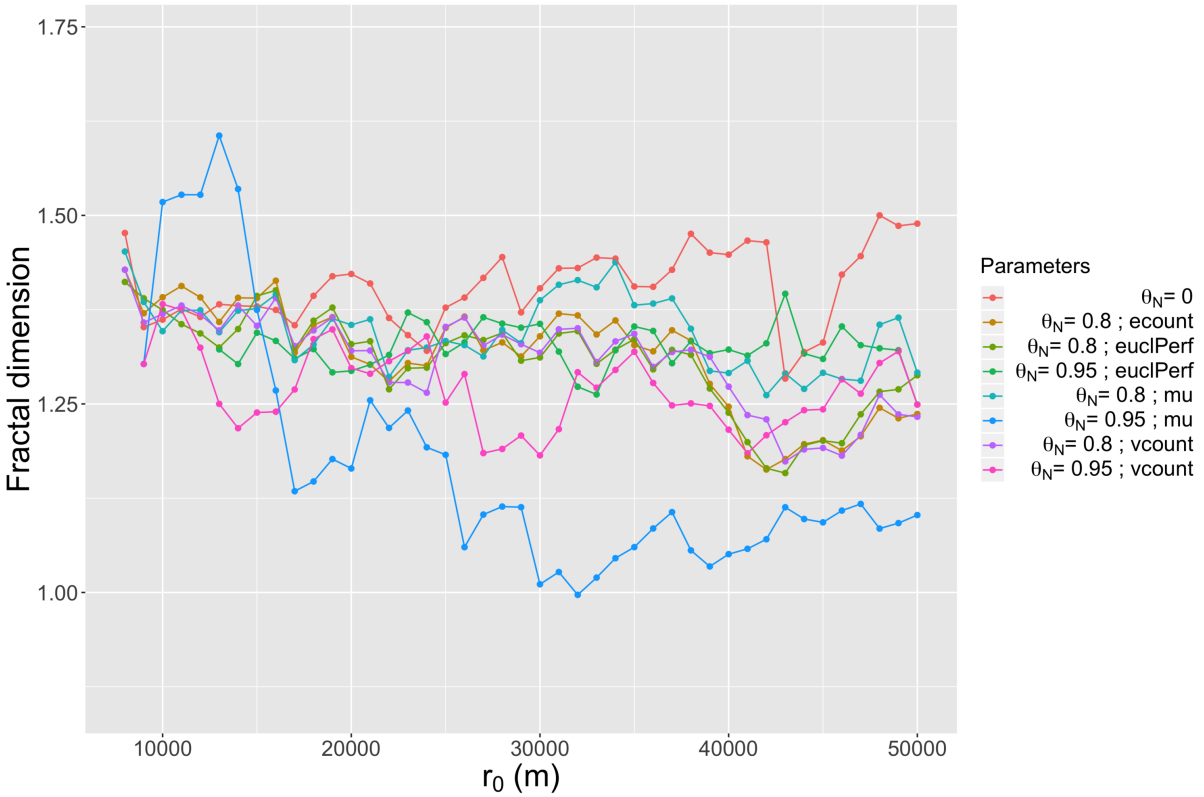


Figure 3: Fractal dimension. We plot for each parametrization given by the curve color the evolution of the fractal dimension α as a function of r_0. Standard errors are not plotted for readability.

The absolute and relative sizes of the largest cluster are plotted in Figure 2 as a function of the percolation radius. This aspect first gives methodological information on multilayer percolation. Indeed, comparing the result with $\theta_N = 0$ (single layer percolation) with positive values of θ_N shows a significantly different behavior. As expected, absolute size are much smaller, but when looking at relative sizes we observe that the abrupt steps typical to percolation transitions have different distributions across the different parametrizations. This result is particularly interesting regarding the first axis of our research question, as it shows that the structure of clusters obtained is not only due to the population layer, and that the multi-dimensional percolation

captures a complementary signal.

The more regular curve seems to be the standard percolation on population only, whereas at $\theta_N = 0.95$, different road network indicators produce either very early transitions (for μ for example) or very late (for N_V). Also, changing of scale compared to Arcaute *et al.* (2016) gives more steps and less abrupts curves in general, confirming the integration of subsystems with different structures in our analysis and the importance of scale in such analysis. As the addition of a layer also changes drastically the results, one should stay careful when switching from a mono-dimensional percolation to a multi-dimensional percolation.

We study also the evolution of the fractal dimension of clusters as a function of r_0, to verify how the initial percolation approach is robust to multi-dimensionality. Following Arcaute *et al.* (2016), we estimate the fractal dimension α of clusters with a simple OLS regression between cluster size and cluster diameter, namely

$$\log N_c = k + \alpha \cdot \log \delta_c \tag{6}$$

where N_c is the size of cluster c and δ_c its diameter. These estimations are shown in Figure 3. As a negative result, which could be due to the abstract nature of our network, a clear maximum in the value of the fractal dimension can not be found. Either it is located at a resolution that our method can not reach due to the minimal 5km limit imposed by the abstraction in the network construction, or it does not exist when coupling dimensions. Determining which assumption is more plausible is out of the scope of this paper.

We do not plot the standard error σ of fractal dimensions (obtained as the estimation error in the OLS) for visibility purposes, but their relative value given by $\alpha/\sigma\left[\alpha\right]$ is in average 0.10 and in maximum 0.196 on all points, meaning that these estimations remain however consistent.

Regarding the variability of fractal dimension as a function of the percolation radius r_0, we study the possible existence of a significant maximum when r_0 varies. We therefore simply consider the difference $(\alpha - \sigma\left[\alpha\right])_M - (\alpha - \sigma\left[\alpha\right])_m$ where the first is taken at the maximum value of α and the other at its minimum value. This intuitively corresponds to checking if confidence intervals do not overlap between the maximum and the minimum of the curve. We find that the configuration for μ and $\theta_N = 0.95$ has a clearly significant maximum (difference at 0.38). For this coupling, the endogenous structure given by the maximum may be defined. Other configurations yield non-significant maximums (negative values of the difference).

This study of percolation transition and fractal dimension thus shows that our multi-dimensional percolation heuristic remains relevant, as results analog but qualitatively different to the one-dimensional approach can be obtained.

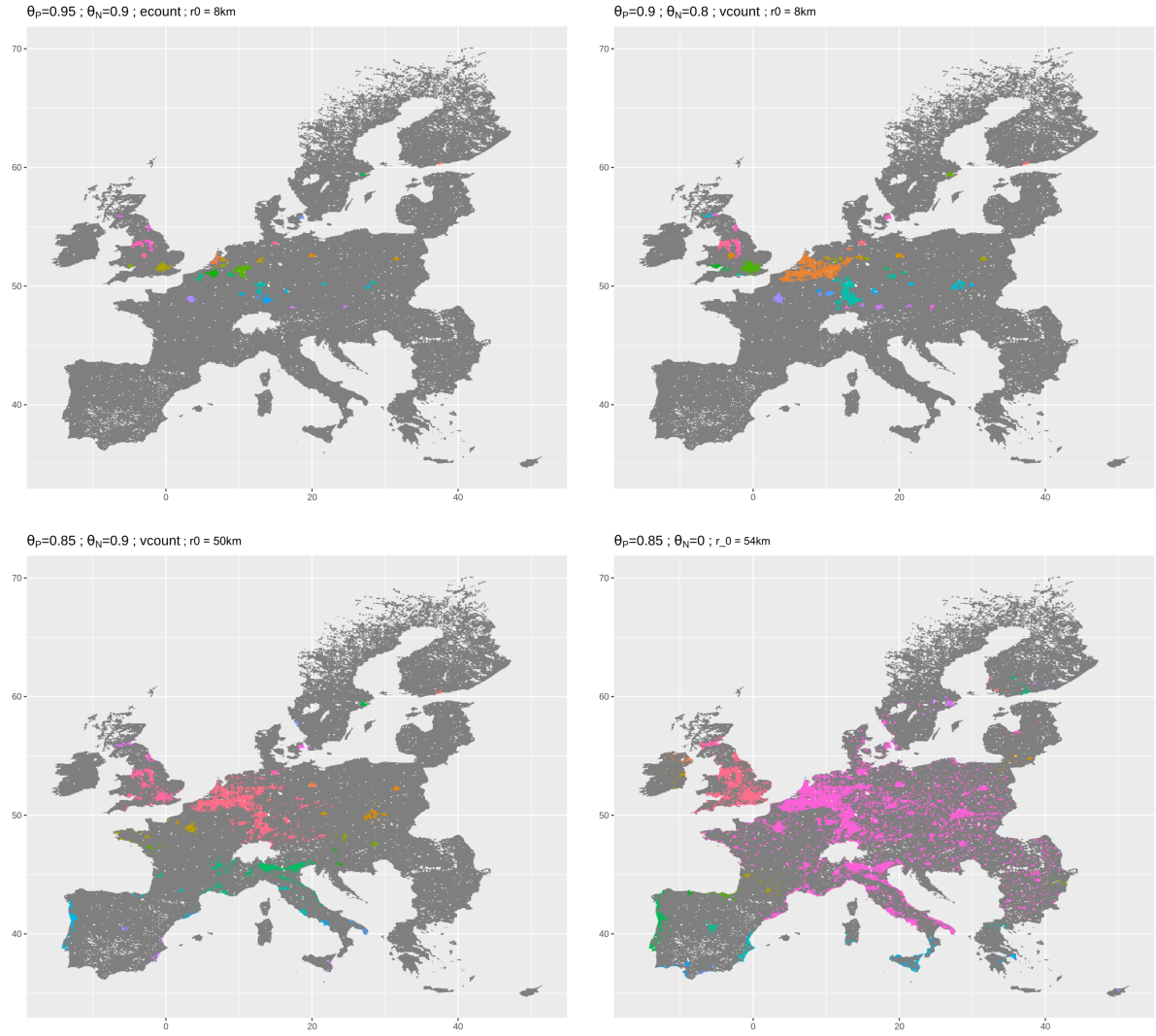


Figure 4: Examples of obtained clusters for different parameter values. In the top-right case for example ($\theta_P = 0.9$, $\theta_N = 0.8$, variable `vcount`,$r_0 = 8km$), we obtain the urban regions of West midlands and London in the UK, Randstad merged with Rhein-Rhur and Rhein-Main in Germany, Paris in France, also with capital cities such as Copenhaguen, Stockholm and Helsinki. There is no cluster in South Europe in that case, due to the high population density threshold.

3.3 Extracting endogenous mega-city regions

We now switch the experience plan to a full grid, for parameters r_0, θ_P, θ_N and the road network indicator considered, and also make γ and d_0 vary. We systematically explore the clusters obtained for 4800 parameter configurations, such that for all road network indicator, $\theta_P \in \{0.8; 0.9; 0.95\}$, $\theta_N \in \{0; 0.8; 0.95\}$, $r_0 \in \{8; 10; 15; 20; 50\}$ km, $\gamma \in \{0.5; 1; 1.5; 2\}$, and $d_0 \in \{0.1; 1; 10; 50; 100\}$ km.

We obtain very different endogenous morphologies for the different parametrizations. Maps reveal that some configurations resemble the actual distribution of European mega-city regions, which are functionally integrated polycentric urban areas (Hall and Pain, 2006). These are here defined endogenously from the bottom-up and have a priori no reason to coincide with these functional regions. We show some examples in Figure 4. The first map of this figure, obtained for high population and network thresholds ($\theta_P = 0.95$ and $\theta_N = 0.9$), but a low radius $r_0 = 8$km and edge count N_E to define the road network layer, include several mega-city

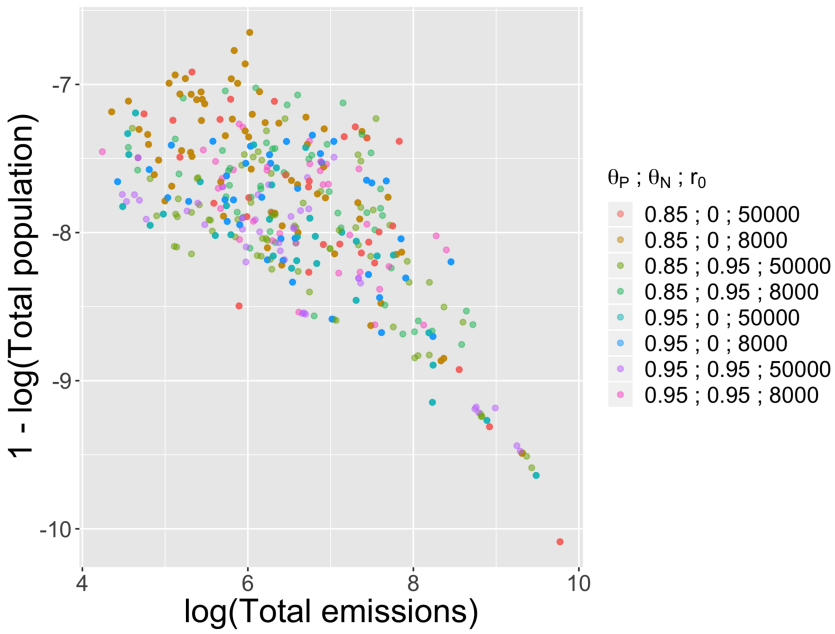


Figure 5: Point clouds of region-level indicators, namely population and emissions, for different parametrizations, given by the color. Each point represent an endogenous urban region.

regions described by Hall and Pain (2006), namely London metropolitan area, the Randstad in Netherland, the Rhein-Main and Rhein-Ruhr in Germany, Greater Paris in France, Brussels area in Belgium. The same parameters with $\theta_N = 0$ yield not exactly the same regions, as confirmed by the transition curves in Figure 2, what means that our approach taking into account two dimensions may capture effective processes of mega-city regions, in particular by including the road network which is crucial as these regions are integrated in terms of flows. The bottom-left map show an example of large clusters emerging in UK and in the center of Europe, the South remaining largely disconnected. Finally, the last map shows the result obtained with a very high radius $r_0 = 54$km, with a giant cluster spanning most of Europe. UK is still disconnected and the transition where it connects happens at $r_0 = 55$km. This does not necessarily mean that UK should be disconnected from continental Europe, as we considered geographic distances only, hiding the high speed connection of the Channel tunnel.

The behavior of sustainability indicators for different population, network and distance thresholds yield different distributions of performances across clusters within a configuration but also between configurations. Before considering the flow-based indicators described above, we can already study basic summary measures such as population P_c and effective emissions EM_c. We show in Figure 5 point clouds of $\log EM_c$ against $1 - \log P_c$ for some configurations. Indeed, regarding the population it contains, an area can be more or less efficient in terms of emissions. Seeing the population as an objective to be maximized (thus the plotted value to be minimized) while the emissions must be minimized, we observe a Pareto front for all points (i.e. all clusters across all configurations). Given different dimensions to be minimized, a Pareto front consists of the points which are not Pareto-dominated by any other point, i.e. that there exists no other point performing best on all objectives. In practice, this yields optimization compromises in the context of multi-objective optimization, when no aggregation of the dimensions is possible or desirable. In Figure 5, the front is the lower bound of the point cloud. We also find no dominating point for each configuration, i.e. that considering point clouds of a single color, a Pareto front with more than one point is still present. Some clusters are therefore optimal compromises

in the Pareto sense in each configuration, while some are dominated and thus not efficient. This confirms that urban systems are generally compromises between multiple objectives.

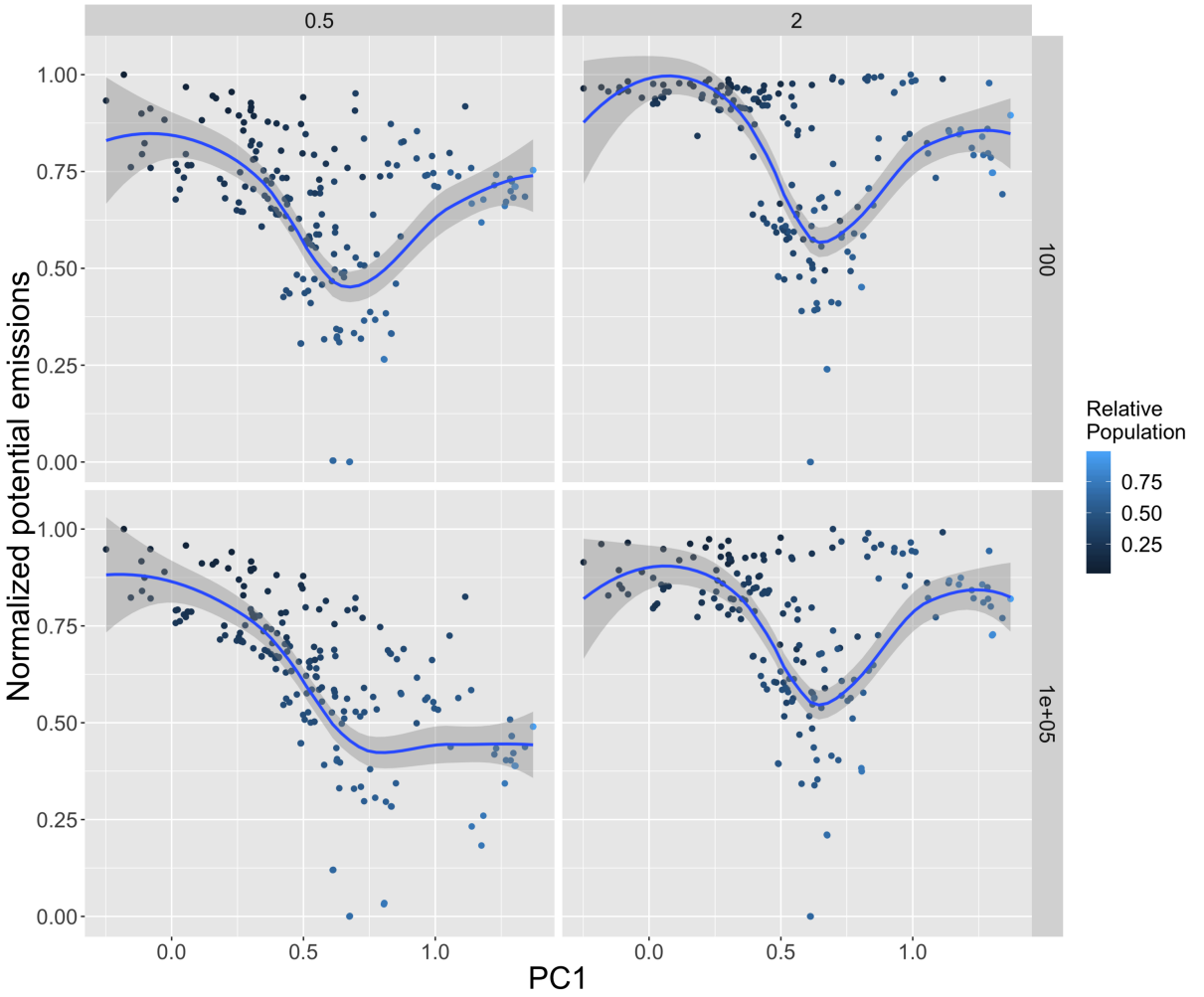


Figure 6: Aggregated values of normalized potential emissions $\sum_c \tilde{g}_c$, as a function of the first morphological principal component (PC1), for varying values of parameters d_G (rows) and γ_G (columns). Other intermediate values for these parameters yield similar behaviors. As PC1 is mainly linked to monocentricity, there seems to exist an optimal intermediate level of monocentricity for emissions alone. Color level give the share of population within the considered clusters in comparison to all European population.

3.4 Linking urban morphology and sustainability

We now consider the sustainability indicators, aggregated for a configuration on all clusters. An important question is how these relate with measures of urban form (Le Néchet, 2012). For a given parametrization of endogenous city regions, one can relate them to morphological indicators for population density spatial distribution. Such indicators were computed by Raimbault (2018a), and capture different dimensions of the spatial distribution of population, such as spatial autocorrelation (Moran index), homogeneity (entropy index), hierarchy (rank-size exponent), or level of aggregation (average distance between individuals). We average them here on clusters. This establishes a link between urban morphology and sustainibility. A principal component analysis on considered points yield 96% of variance with two components, and 73% explained by the first component alone. The first component relates to a level of monocentricity ($PC1 = -0.3 \cdot I + 0.54 \cdot \bar{d} + 0.51 \cdot \varepsilon + 0.59 \cdot h$ where I is Moran index, $\bar{d}$ average distance,

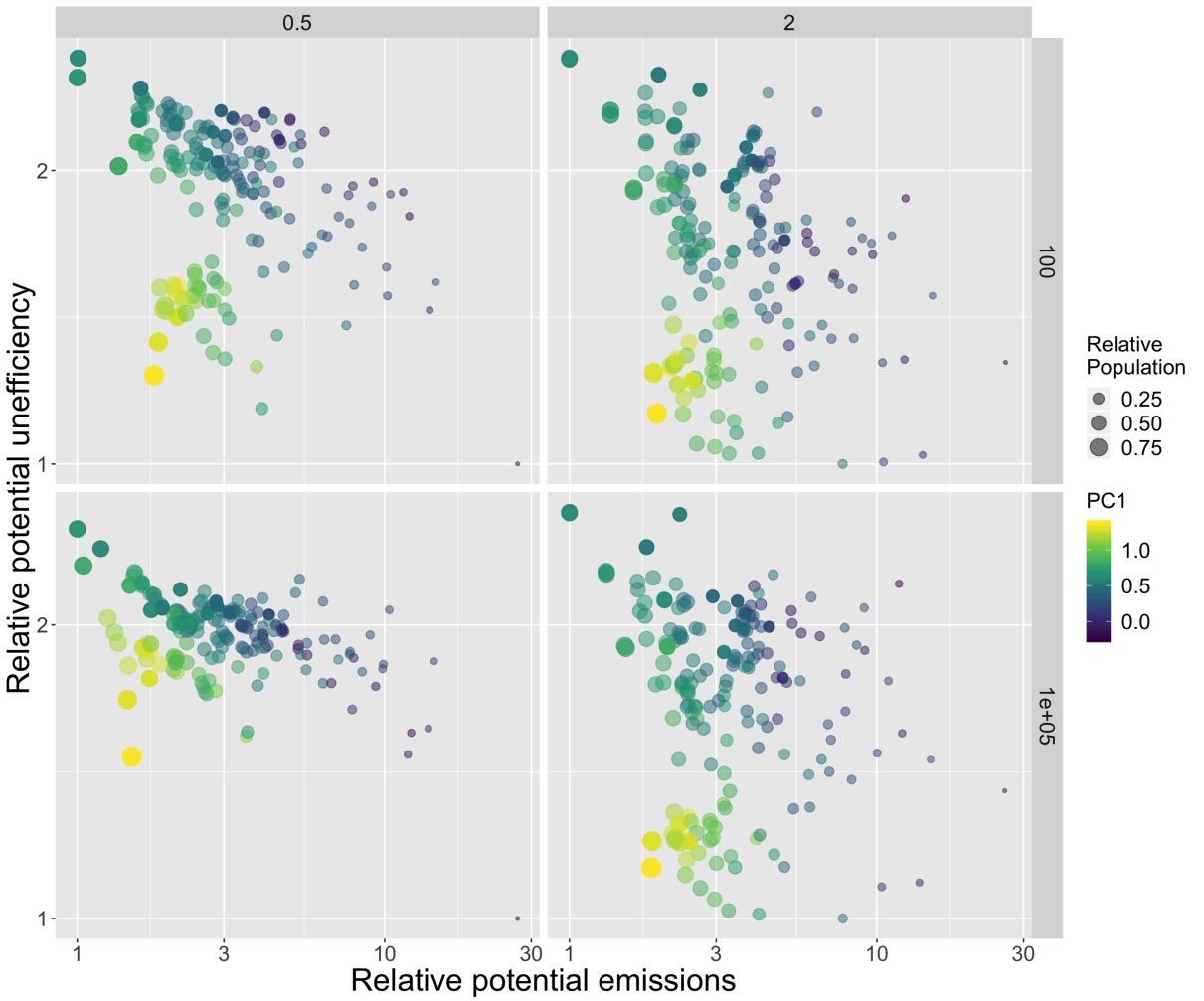


Figure 7: Relative potential emissions $\sum_c g_c$ against relative potential economic inefficiency $\sum_c e_c$ (both indicators should be minimized), for varying values of γ_G (columns) and d_G (rows). Color level gives the value of PC1, whereas point size gives the share of total population contained within considered areas.

ϵ entropy, and h level of hierarchy). In a nutshell, the principal dimension of urban form in differentiating our urban clusters is the level of monocentricity. We can relate it with indicators for emissions and economic efficiency.

We show in Figure 6 the value of $\sum_c \tilde{g}_c$ as a function of the first morphological principal component, for extreme values of γ and d_0. There seems to exist an optimal intermediate value for PC1 regarding the minimization of the normalized indicator for emissions only. This would correspond to an intermediate level of monocentricity, meaning that urban areas which are too polycentric and spread would emit more, but also areas that are too much monocentric. This behavior does not occur for long-range $d_0 = 100$km and low-hierarchy $\gamma = 0.5$ interactions. The mostly monocentric but emitting configurations capture most of population (given by the level of color), whereas the intermediate configurations capture around half of the population, what means that these low-emissions potential urban regions can cover a significant part of European population.

However, when considering both emissions and economic indicators, urban form then acts as a compromise variable. We show in Figure 7 the point clouds of $\sum_c g_c$ against $\sum_c e_c$, which produce clear Pareto fronts, which shape varies with γ and d_0. As the color level gives the value of

PC1, we can see the points on the different fronts with very different morphological properties. In some cases, highly monocentric areas (yellow points) can be a good compromise, whereas the intermediate optimal for emissions shown before may yield highly inefficient areas (dominated green points). For example, considering the fronts for $\gamma = 2$ which have both very similar shape, the points with the lowest emissions are on the top-left of the front and correspond to the optimal unveiled in Figure 6. These have however a very low economic efficiency (high inefficiency) and small improvements can be done with the points below, before switching to a totally different urban form with a high value of PC1 (yellow points, highly monocentric). Increasing more the economic efficiency is then at the price of much more emissions, with more polycentric areas. This analysis therefore unveils morphological trade-offs, confirming that there is no optimal urban form, but different compromises regarding the conflicting sustainability indicators.

IV DISCUSSION

4.1 From multi-dimensional percolation to the sustainability of mega-city regions

From the methodological point of view, we showed that network percolation can successfully be applied to multidimensional urban networks. This requires a consistent overlay within the same nodes of the different dimensions considered. The existence of percolation transitions which are different to the unidimensional case confirms that the approach captures complementary information, and that it could be applied to characterize urban systems in a more refined way. The non-existence of a significant maximum for the fractal dimension remains to be investigated, since it could be due a bias of our abstract network construction. Studies on other dimensions or on non-abstract networks remain to be done to understand how multi-dimensional percolation differs from the unidimensional one.

Regarding the second axis of our research question, we showed that multi-dimensional percolation is a useful tool to extract endogenous mega-urban regions while taking into account complementary aspects of population distribution and performance of transportation networks. Varying the parameters of the percolation algorithm provides a comparative view on possible clustering structures for the European urban system, and corresponding performance in terms of stylized sustainability indicators. Indeed, this work is exploratory in terms of possible definitions of urban subsystems. The fact that some systems obtained coincide with effective functional regions (Hall and Pain, 2006) shows that some thresholds of population and road network performance intensity capture actual functional linkages. This correspondance could not be predicted a priori nor explained through simple arguments as our approach reconstructs clusters from the bottom-up. Finally, the links between urban form and sustainability indicators made in the last section are also interesting for the management of urban systems, suggesting a certain performance of polycentric systems in particular regarding emissions.

4.2 Developments

Further work may consist in the use of calibration heuristics to find in a more robust way optimal parameter values. The OpenMOLE model exploration platform provides a transparent access to genetic algorithms for multi-objective optimization (Reuillon *et al.*, 2013). The use of such calibration algorithms would allow unveiling the effective form of Pareto fronts, that we may have missed here through the grid sampling.

An other development would consist in extrapolating transportation flows with a spatially explicit gravity and transportation flow model as a kind of simplified four step model (McNally,

2000). It could then be adjusted on actual transportation flows emissions database which are also available in the Edgar database. The corresponding gravity parameters could then be used within the economic and emissions potentials, and the sustainability patterns produced compared with the hypothetical ones we produced here.

Finally, an important development would imply crossing our endogenous definitions of urban regions with socio-economic databases, and compute indicators implied in other dimensions of sustainability, for example related to socio-economic inequalities, spatial distribution of accessibilities, or activities with different scaling exponents. This includes the mitigation of spatial inequalities and segregation (Tammaru *et al.*, 2015), which are an important dimension of sustainability.

4.3 Towards policy applications

Our work suggests the possibility to design policies in terms of regional integration to increase the sustainability of mega-city regions. The way such results could actually be transferred to policy-making recommandations remains an open question, but Pareto-optimal configurations can be used for the planning of regional transportation networks for example, or to design policies for the distribution of subsidies. Indeed, privileging some infrastructure developments but also collaborations between urban centers can be seen as an aspect of a small scale planning, or territorial strategy. As we integrated potential flows that would result from such development, and considered their economic and emissions consequences, and did it in an endogenous way, we suggest that evidence-based strategies for territorial development at the European level could be inspired by this work. This would naturally imply a more thorough data integration, model calibration and operationalization.

V CONCLUSION

In conclusion, our multilayer percolation approach captures in a way the multi-dimensionality of urban systems and a link between form and function in urban system. In particular, in our application on the bilayer case of an abstract network constructed from population density and road network indicators, it is shown to capture a different structure than in the unidimensional case. Its application to the issue of sustainable mega-city regions show how its properties of constructing urban clusters from the bottom-up can be used to study sustainability issues. This work also illustrates the importance of following data-driven paradigms even when developing new methods, as what is understood of the behavior of the heuristic is through its application to real data and issues.

ACKNOWLEDGMENT

This work is part of DynamiCity, a FUI project funded by BPI France, Auvergne-Rhône-Alpes region, Ile-de-France region and Lyon metropolis.

References

Aleta A., Y. M. (2018). Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics 10*, 45–62. doi:10.1146/annurev-conmatphys-031218-013259.

Arbabi H., Mayfield M., McCann P. (2019). On the development logic of city-regions: inter-versus intra-city mobility in England and Wales. *Spatial Economic Analysis 14*(3), 1–20. doi:10.1080/17421772.2019.1569762.

Arcaute E., Molinero C., Hatna E., Murcio R., Vargas-Ruiz C., Masucci A. P., Batty M. (2016). Cities and regions in Britain through hierarchical percolation. *Royal Society open science 3*(4), 150691. doi:10.1098/rsos.150691.

Banos A., Genre-Grandpierre C. (2012). Towards new metrics for urban road networks: Some preliminary evidence from agent-based simulations. In *Agent-based models of geographical systems*, pp. 627–641. Springer. doi:10.1007/978-90-481-8927-4_31.

Barthélemy M. (2011). Spatial networks. *Physics Reports 499*(1-3), 1–101. doi:10.1016/j.physrep.2010.11.002.

Batty M., Longley P. A. (1994). *Fractal cities: a geometry of form and function*. Academic press. URL: https://dl.acm.org/citation.cfm?id=179240.

Bettencourt L. M., Lobo J., Helbing D., Kühnert C., West G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences 104*(17), 7301–7306. doi:10.1073/pnas.0610172104.

Boccaletti S., Bianconi G., Criado R., Del Genio C. I., Gómez-Gardenes J., Romance M., Sendina-Nadal I., Wang Z., Zanin M. (2014). The structure and dynamics of multilayer networks. *Physics Reports 544*(1), 1–122. doi:10.1016/j.physrep.2014.07.001.

Burger M., Meijers E. (2012). Form follows function? linking morphological and functional polycentricity. *Urban studies 49*(5), 1127–1149. doi:10.1177/0042098011407095.

Callaway D. S., Newman M. E., Strogatz S. H., Watts D. J. (2000). Network robustness and fragility: Percolation on random graphs. *Physical review letters 85*(25), 5468. doi:10.1103/PhysRevLett.85.5468.

Cottineau C., Finance O., Hatna E., Arcaute E., Batty M. (2018). Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science in press*. doi:10.1177/2399808318755146.

Cottineau C., Hatna E., Arcaute E., Batty M. (2017). Diverse cities or the systematic paradox of urban scaling laws. *Computers, environment and urban systems 63*, 80–94. doi:10.1016/j.compenvurbsys.2016.04.006.

Crucitti P., Latora V., Porta S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E 73*(3), 036125. doi:10.1103/PhysRevE.73.036125.

Ducruet C., Beauguitte L. (2014). Spatial science and network science: Review and outcomes of a complex relationship. *Networks and Spatial Economics 14*(3-4), 297–316. doi:10.1007/s11067-013-9222-6.

Feng Y., Wu S., Wu P., Su S., Weng M., Bian M. (2018). Spatiotemporal characterization of megaregional polycentrality: Evidence for new urban hypotheses and implications for polycentric policies. *Land Use Policy 77*, 712–731. doi:10.1016/j.landusepol.2018.06.022.

Florida R., Gulden T., Mellander C. (2008). The rise of the mega-region. *Cambridge Journal of Regions, Economy and Society 1*(3), 459–476. doi:10.1093/cjres/rsn018.

Hackett A., Cellai D., Gómez S., Arenas A., Gleeson J. P. (2016). Bond percolation on multiplex networks. *Physical Review X 6*(2), 021002. doi:10.1103/PhysRevX.6.021002.

Hall P. G., Pain K. (2006). *The polycentric metropolis: learning from mega-city regions in Europe*. Earthscan. URL: http://www.strategvest.ro/media/dms/file/Resurse/The%20Polycentric%20Metropolis%202006.pdf.

Hillier B., Leaman A., Stansall P., Bedford M. (1976). Space syntax. *Environment and Planning B: Planning and design 3*(2), 147–185. doi:10.1068/b030147.

Huynh H. N., Makarov E., Legara E. F., Monterola C., Chew L. Y. (2018). Characterisation and comparison of spatial patterns in urban systems: A case study of U.S. cities. *Journal of Computational Science 24*, 34–43. doi:10.1016/j.jocs.2017.12.001.

Janssens-Maenhout G., Crippa M., Guizzardi D., Muntean M., Schaaf E., Dentener F., Bergamaschi P., Pagliari V., Olivier J., Peters J., van Aardenne J. A., Monni S., Doering U., Petrescu A. M. R. (2017). EDGAR v4.3.2 global atlas of the three major greenhouse gas emissions for the period 1970–2012. *Earth System Science Data 11*, 959–1002. doi:10.5194/essd-11-959-2019.

Komiyama H., Takeuchi K. (2006). Sustainability science: building a new discipline. *Sustainability Science 1*(1), 1–6. doi:10.1007/s11625-006-0007-4.

Lagesse C., Bordin P., Douady S. (2015). A spatial multi-scale object to analyze road networks. *Network Science 3*(1), 156–181. doi:10.1017/nws.2015.4.

Lang R. E., Dhavale D. (2005). Beyond megalopolis: Exploring america's new "megapolitan" geography.

Technical Report 38, Brookings Mountain West. URL: https://digitalscholarship.unlv.edu/brookings_pubs/38/.

Laquian A. (2011). The planning and governance of Asia's mega-urban regions. In *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*, Chapter 14, pp. 302–322. United Nations Publication. URL: https://drive.uqu.edu.sa/_/aeelfarouk/files/PopulationDistributionUrbanization.pdf.

Lashof D. A., Ahuja D. R. (1990). Relative contributions of greenhouse gas emissions to global warming. *Nature 344*, 529–531. doi:10.1038/344529a0.

Le Néchet F. (2012). Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities. *Cybergeo: European Journal of Geography*, 580. doi:10.4000/cybergeo.24966.

Le Néchet F. (2017). De l'étalement urbain aux régions métropolitaines polycentriques : formes de fonctionnement et formes de gouvernance. In *Peupler la terre - De la préhistoire à l'ère des métropoles*, Chapter 13, pp. 361–383. Presses Universitaires Francois Rabelais. doi:10.4000/books.pufr.10467.

Li D., Fu B., Wang Y., Lu G., Berezin Y., Stanley H. E., Havlin S. (2015). Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences 112*(3), 669–672. doi:10.1073/pnas.1419185112.

Louf R., Barthelemy M. (2013). Modeling the polycentric transition of cities. *Physical review letters 111*(19), 198702. doi:10.1103/PhysRevLett.111.198702.

Makse H. A., Andrade J. S., Batty M., Havlin S., Stanley H. E. (1998). Modeling urban growth patterns with correlated percolation. *Physical Review E 58*(6), 7054. doi:10.1103/PhysRevE.58.7054.

Marull J., Galletto V., Domene E., Trullén J. (2013). Emerging megaregions: A new spatial scale to explore urban sustainability. *Land Use Policy 34*, 353–366. doi:10.1016/j.landusepol.2013.04.008.

McNally M. G. (2000). The four step model. Technical Report UCI-ITS-AS-WP-07-2, University of California Irvine, Institute of Transportation Studies, Center for Activity Systems Analysis. URL: https://escholarship.org/uc/item/0r75311t.

Newman M. E. J., Watts D. J. (1999). Scaling and percolation in the small-world network model. *Physical Review E 60*(6), 7332–7342. doi:10.1103/PhysRevE.60.7332.

Offner J.-M., Beaucire F., Delaplace M., Frémont A., Ninot O., Bretagnolle A., Pumain D. (2014). Les effets structurants des infrastructures de transport. *Espace Geographique 43*, 51–67. doi:10.3917/eg.431.0051.

Penrose M. D. (1999). On k-connectivity for a geometric random graph. *Random Structures & Algorithms 15*(2), 145–164. doi:10.1002/(SICI)1098-2418(199909)15:2¡145::AID-RSA2¿3.0.CO;2-G.

Perez P., Banos A., Pettit C. (2016). Agent-based modelling for urban planning current limitations and future trends. In *International Workshop on Agent Based Modelling of Urban Systems*, Volume 10051 of *Lecture Notes in Computer Science*, pp. 60–69. doi:10.1007/978-3-319-51957-9_4.

Piovani D., Molinero C., Wilson A. (2017). Urban retail location: insights from percolation theory and spatial interaction modeling. *PloS One 12*(10), e0185787. doi:10.1371/journal.pone.0185787.

Raimbault J. (2018a). Calibration of a density-based model of urban morphogenesis. *PloS one 13*(9), e0203516. doi:10.1371/journal.pone.0203516.

Raimbault J. (2018b). *Caractérisation et modélisation de la co-évolution des réseaux de transport et des territoires*. Phd thesis, Université Paris 7 Denis Diderot. URL: https://tel.archives-ouvertes.fr/tel-01857741.

Raimbault J. (2018c). Indirect evidence of network effects in a system of cities. *Environment and Planning B: Urban Analytics and City Science in press*. doi:10.1177/2399808318774335.

Raimbault J. (2019). An urban morphogenesis model capturing interactions between networks and territories. In *The Mathematics of Urban Morphology*, Modeling and Simulation in Science, Engineering and Technology, pp. 383–409. Springer. doi:10.1007/978-3-030-12381-9_17.

Reuillon R., Leclaire M., Rey-Coyrehourcq S. (2013). OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems 29*(8), 1981–1990. doi:10.1016/j.future.2013.05.003.

Sanders L. (2017). *Peupler la terre - De la préhistoire à l'ère des métropoles*. Villes et territoires. Presses Universitaires Francois Rabelais. URL: https://pufr-editions.fr/produit/peupler-la-terre/.

Son S.-W., Bizhani G., Christensen C., Grassberger P., Paczuski M. (2012). Percolation theory on interdependent networks based on epidemic spreading. *Europhysics Letters 97*(1), 16006. doi:10.1209/0295-5075/97/16006.

Sorensen A., Okata J. (2010). *Megacities: Urban form, governance, and sustainability*, Volume 10 of *cSUR-UT Series: Library for Sustainable Urban Regeneration*. Springer. doi:10.1007/978-4-431-99267-7.

Stauffer D., Aharony A. (2014). *Introduction to percolation theory*. CRC Press / Taylor & Francis. URL: https://www.crcpress.com/Introduction-To-Percolation-Theory-Second-Edition/Stauffer-Aharony/p/book/9780748402533.

Su S., Liu Z., Xu Y., Li J., Pi J., Weng M. (2017). China's megaregion policy: Performance evaluation framework, empirical findings and implications for spatial polycentric governance. *Land Use Policy 63*, 1–19. doi:10.1016/j.landusepol.2017.01.014.

Tammaru T., Marcińczak S., Van Ham M., Musterd S. (2015). A multi-factor approach to understanding socio-economic segregation in European capital cities. In *Socio-economic segregation in European capital cities - East Meets West*, Chapter 1, pp. 1–29. Taylor & Francis. doi:10.4324/9781315758879.

Tobler W. (2004). On the first law of geography: A reply. *Annals of the Association of American Geographers 94*(2), 304–310. doi:10.1111/j.1467-8306.2004.09402009.x.

Viguié V., Hallegatte S. (2012). Trade-offs and synergies in urban climate policies. *Nature Climate Change 2*, 334–337. doi:10.1038/nclimate1434.

Zeng G., Li D., Guo S., Gao L., Gao Z., Stanley H. E., Havlin S. (2019). Switch between critical percolation modes in city traffic dynamics. *Proceedings of the National Academy of Sciences 116*(1), 23–28. doi:10.1073/pnas.1801545116.