Spatially-encouraged spectral clustering:

a technique for blending map typologies and regionalization

Levi John Wolfa

aSchool of Geographical Sciences, University of Bristol, United Kingdom.

**ARTICLE HISTORY**

Compiled May 19, 2021

**ABSTRACT**

Clustering is a central concern in geographic data science and reflects a large, active domain of research. In spatial clustering, it is often challenging to balance two kinds of “goodness of fit:” clusters should have “feature” homogeneity, in that they aim to represent one “type” of observation, and also “geographic” coherence, in that they aim to represent some detected geographical “place.” This divides “map ty- pologization” studies, common in geodemographics, from “regionalization” studies, common in spatial optimization and statistics. Recent attempts to simultaneously typologize and regionalize data into clusters with both feature homogeneity and geo- graphic coherence have faced conceptual and computational challenges. Fortunately, new work on spectral clustering can address both regionalization and typologization tasks within the same framework. This research develops a novel kernel combination method for use within spectral clustering that allows analysts to blend smoothly be- tween feature homogeneity and geographic coherence. I explore the formal properties of two kernel combination methods and recommend multiplicative kernel combina- tion with spectral clustering. Altogether, spatially-encouraged spectral clustering is shown as a novel kernel combination clustering method that can address both re- gionalization and typologization tasks in order to reveal the geographies latent in spatially-structured data.

**KEYWORDS**

clustering; geodemographics; spatial analysis; spectral clustering

# Introduction

Clustering is a critical concern in geography. In geographic data science, *unsupervised learning* has emerged as a remarkably effective collection of approaches for describ- ing and exploring the latent structure in data (Hastie et al. 2009, Grubesic et al. 2014). Spatial clustering has many common applications in epidemiology (Turnbull et al. 1990, Besag and Newell 1991, Kulldorff and Nagarwalla 1995, Neill et al. 2005, Rogerson and Yamada 2009) and econometrics (Czamanski and Ablas 1979, Rey and Mattheis 2000, Arbia et al. 2008). Further, the exploratory analysis of spatial outliers and “hotspots” is a common technique in exploratory spatial data analysis (Anselin 1995, Getis and Ord 1996), and is a common mode of analysis for an initial geographic interrogation. The determination of spatially meaningful communities also plays a large role in urban sociology (Galster 2001, Drukker et al. 2003, Spielman and Logan

Corresponding Author: Levi John Wolf. Email: [levi.john.wolf@bristol.ac.uk](mailto:levi.john.wolf@bristol.ac.uk)

2013), and the visualization of geography inherent in demographic data is a robust field known as geodemographics (Harris et al. 2005, 2007, Singleton and Longley 2009, Singleton and Spielman 2014). Since the regions identified in these methods are never “known” (and indeed, may be difficult *to* know (Isard 1956)), the identification and de- lineation of geographic clusters or regional partitions stands as an enduring analytical need.

In a geographical problem, clustering is often used to identify distinctive “formal re- gions” (Shelton and Poorthuis 2019) or to uncover “the geography of specific domains” in a given social process (Singleton and Spielman 2014).1 These two motives for cluster- ing entail different representations of similarity, despite their shared academic heritage (Knaap et al. 2019). As both Shelton and Poorthuis (2019) and Singleton and Spiel- man (2014) discuss, the notion of “places that are uniform and homogeneous, grouped together based on their similarity according to a series of given variables” (Shelton and Poorthuis 2019, p.1345) has “its origins in the work of human ecologists in the 1920s and 1930s.” (Singleton and Spielman 2014, p. 559).

As with any salient division in social science literature, the distinctions drawn be- tween the two bodies of literature stem from the stage of analysis into which geogra- phy enters. *Regionalization* focuses on identifying geographically-coherent and feature- homogeneous “formal regions,” so geography is an exogenously-determined “input” to analysis. *Map Typologization*, typical in geodemographics (Harris et al. 2005) or tra- jectory analysis (Delmelle et al. 2013, Zwiers et al. 25 2016), usually ignores geograph- ical coherence at the outset, but then explores the geography of uncovered solutions. Unfortunately, “[u]nless the clustering technique is explicitly spatial, the geographic relevance [of clusters] might not be sufficiently accounted for.” (Grubesic et al. 2014), so it may be difficult to interpret geodemographic classifications after the fact. How- ever, geodemographic analyses are used precisely because interpretable geographies often emerge from geographical data.

For methods that explicitly seek to mediate this divide, it can be difficult to ensure that clusters both represent a place as well as a type of observation (Zhu et al. 2020). Therefore, “geographic coherence,” the idea that a cluster is relatively compact and identifies a “region,” is often in tension with the internal self-similarity of clusters according to their data, called “feature homogeneity” (Wolf et al. 2019). In fact, this is often because “geographic coherence” is modeled as a binary process, where regions are considered “coherent” if and only if they are geographically contiguous. Other ways to represent geographic coherence, such as Wu and Murray (2008), offer alternatives that are not binary, but continuous.

In this vein, affinity clustering provides a flexible way to blend geographic coher- ence and feature homogeneity. In this paper, I propose an alternative method to blend *regionalization* and *typologization* using a kernel combination. I demonstrate the effec- tiveness of a kernel combination using an affinity-based clustering technique: spectral clustering. However, the kernel combination concept shown here can be applied to a wide variety of affinity-based clustering algorithms.

To explain the method in full, I outline recent advancements in spectral clustering and pay special attention to how spectral clustering can be distinctly useful for ge- ographical problems. I discuss recent work using spectral clustering for geographical processes, and highlight a new kernel combination approach. I discuss how using a

1Further types of “core detection” (Aldstadt and Getis 2006, Murray et al. 2014, Kim et al. 2017) or “boundary detection” (Jacquez et al. 2008, Dean et al. 2018, Dong et al. 2018) allow for “non-exhaustive” partitions, where observations can evade cluster assignments. This is not of interest at here—using Kim et al. (2017)’s terminology, this means only “districting” methods are considered.

kernel combination within spectral clustering can satisfy both “strict” geographical constraints (akin to those used in mixed-integer programming regionalization (Duque et al. 2011, Oehrlein and Haunert 2017)) and “soft” constraints that encourage the solution towards geographical coherence. I examine how this approach differs from past approaches, both mathematically and empirically using two examples: clustering presidential vote share swings in Texas counties and clustering Airbnb price data in Brooklyn, NY.

# Methods

Spectral clustering provides an opportunity to construct unsupervised learning so- lutions that satisfy needs of both regionalization and typologization tasks. Spectral clustering is commonly used in machine learning to analyze the latent spatial struc- ture in data (Shi and Malik 2000, Von Luxburg 2007). As a clustering technique, it has often been applied to graph-embedded data (White and Smyth 2005). The principal mathematical ideas are well known in spatial econometrics (Griffith 2000, 2013) and often find their way into supervised learning of geographic processes. However, spec- tral clustering methods have recently been applied to the analysis of segregation and sorting in American cities (Chodrow 2017), and will become more useful for spatial analysis as their theory and implementations improve.

## *Recent Work on Clustering & Regionalization*

A few recent studies have attempted to relax the strictness of spatial constraints in or- der to blend typologization and regionalization solutions. One recent exploration uses novel spatial methods in an analysis of group separation and segregation (Chodrow 2017). However, Chodrow (2017) does not directly parameterize the balance between geographic coherence and feature homogeneity. This lack of a direct parameterization makes it difficult to assess the trade-off between geographic coherence and feature homogeneity. However, spectral clustering methods (Shi and Malik 2000, Ng et al. 2002, Von Luxburg 2007) can indeed include an explicitly-parameterized trade-off be- tween geographic coherence and feature homogeneity. This is possible because spectral clustering tends to produce relatively-compact regions of similar sizes (Shi and Malik 2000) which are usually of interest for regionalization.

One recent innovation, *spatially-constrained spectral clustering* (Yuan et al. 2015)

provides an explicit trade-off between geographic coherence and feature homogeneity within spectral clustering. The method is novel in part because a single parameter is used to relax geographic coherence constraints. A parameterized trade-off between geographic coherence and feature homogeneity makes it possible to assess the full de- cision frontier between geographic-driven regions and feature-driven clusters. Further, it means that we can consider the balance between geographic coherence and feature homogeneity *while building clusters*, rather than at the end.

To develop the full implications of this innovation, it is useful to provide back- ground on spectral clustering methods. The discussion below contextualizes Yuan et al. (2015)’s innovations within the theory of spectral clustering. Therefore, this discussion highlights the core improvements by Yuan et al. (2015) to constrained spectral cluster- ing. Afterwards, this research explores the impact of an omitted free parameter—the feature kernel bandwidth. While Yuan et al. (2015) suggest that only the spatial band- width parameter controls geographic coherence, this research shows that both feature

and geographic kernels affect solution contiguity in relaxed problems. As such, the findings demonstrate that both feature and geographic kernels must be fully parame- terized, which leads to a more complete statement of spatial spectral clustering.

## *Fundamentals of Spectral Clustering*

Spectral clustering works by finding clusters in a lower-dimensional embedding of high- dimensional feature data (Shi and Malik 2000, Ng et al. 2002). Thus, at a high level, spectral clustering has two steps:

* + 1. **dimension reduction**: relevant eigenvectors are extracted from a summary matrix of the data.
    2. **cluster discovery**: clusters are detected from the lower-dimensional embedding.

The data summary matrix, called the *affinity matrix*, encodes the similarity between each pair of observations’ features. Typically, feature similarity increases as the dis- tance between observations’ feature vectors decreases—observations with similar fea- tures have higher affinity.2 With an affinity matrix, spectral clustering operates on the implied *Laplacian* matrix of this affinity structure, which is defined as:

**L** = **D** *−* **A** (1)

where **A** is the affinity matrix recording the similarity between observations, **D** is

the ”degree” matrix, with **D***ii* = L*N* **A***ij* and zero elsewhere, and **L** is the implied

*j*=1

*Laplacian* matrix. Then, the set of *K* eigenvectors corresponding to the largest *K* eigenvalues extracted from **L** provide a new set of lower-dimensional data over which a simpler method, such as *K*-means or hierarchical partitioning, can be used to find clusters more efficiently (Von Luxburg 2007).

## *Spatial Spectral Clustering using Kernel Combination*

Practically speaking, classic spectral clustering provides an approximate solution to a normalized minimum weighted cut problem in image segmentation (Shi and Malik 2000). As detailed by Shi and Malik (2000), spectral clustering was developed to cut images into patches that “look” distinctive and usually represent entities of interest. In this task, it is desirable to avoid segmenting images into too-small or too-spindly parts, since extremely small patches rarely represent entities in their own right. Thus, spectral clustering (and affiliated methods) find reasonably-large consistently-sized clusters3 in the weighted graph defined by an affinity matrix, **A**, defined between all pairs of pixels.

The same concerns about the size and shape of partitions also occur in regionaliza- tion and clustering. There, clusters usually need to be of a sufficient size and reasonable shape, but these properties are usually not represented with specific parameters. In- terestingly, when **A** contains *only* spatial information (e.g. spatial adjacency or kernel relationships) spectral clustering will tend to find boundaries that cut the map into compact regions of similar size. This is because the method still approximates a so- lution to the normalized minimum weighted cut that partitions the affinity matrix.

2Numerically, it is common to use a *kernel function*, such as the negative exponential kernel, and standardize the resulting values to between 0 and 1.

3although the minimum size or shape regularity are not parameterized directly as in other methods (Duque et al. 2012, Li et al. 2014)

For solely-spatial affinities, this means spectral clustering tends to reduce the weight of the links that cross cluster boundaries while ensuring no area has too small a total weight. Hence, its regions will be compact and regularly-sized.

What Yuan et al. (2015) realize is that this spatial structure can be influenced. To do this, they split their affinity matrix into two parts: **A***f* , a feature affinity matrix and **A***s*, a spatial constraint matrix. For their problem, Yuan et al. (2015) parameterize the spatial affinity matrix using a single parameter, *η*. This *η* parameter is the parameter governing the shape of the spatial kernel, **A***s*, and is the only parameter directly considered by Yuan et al. (2015). Together, the form for **L***s*, the spatially-informed Laplacian that Yuan et al. (2015) use for clustering, is similar to standard spectral clustering:

**L***s* = **D***s −* **A***f ◦* **A***s*(*η*) (2)

where **D***s* again has the row sums of **A***f ◦* **A***s* on its diagonal and zero elsewhere and *◦* denotes the element-wise product of two matrices. This element-wise product filters the feature affinities through spatial constraints, a kind of *kernel combination*. The final Laplacian matrix, **L***s*, represents a graph containing the affinity matrix **A***f* combined with geographic information in **A***s*. Must-link (1) and cannot-link (*−*1) entries in **A***s* can be used as well (Wang and Davidson 2010) which provide the same kinds of contiguity requirements as present in geographical regionalization problems (Duque et al. 2011).

Fortunately, the properties proven by Shi and Malik (2000) for generalized spectral clustering still apply to solutions obtained from combined kernels. As Wang et al. (2014) show, clusters form an approximate solution to the original normalized mini- mum weighted cut problem while obeying new constraints.

# Defining Spatially-Encouraged Spectral Clustering

In fact, the expression used in Eq. 2 has a latent parameter that controls the shape of the feature kernel, **A***f* . Yuan et al. (2015) model the feature similarity between a pair sites (*i* and *j*) using a Gaussian kernel:

[**A***f* ]*ij* = exp *−τ* 2*||***X***i −* **X***j||*2 (3)

where *||***X***i −* **X***j||* denotes the Euclidean distance between the observations’ *P* -length feature vectors. Thus, *τ* 2 is the “decay” parameter of their feature kernel, where *θ* =  1 would describe the feature bandwidth. In this framework, *θ* is the feature kernel’s analogue of the geographic kernel’s spatial bandwidth (*η*): large values of *θ* mean that observations with very different features are considered similar, just as large *η* mean that observations at very different locations are considered near. Yuan et al. (2015) hold *τ* 2 constant at an unspecified value throughout their experiments, as they are only interested in the spatial bandwidth parameter (*η*).4

*τ* 2

4In their specific case, Yuan et al. (2015) cluster principal components derived from many mean-centered and unit-deviation standardized covariates. But, *τ* 2 is not intended to stand in as the empirical variance of **X** generally, as **X** may be *N × P* with different variances for each feature but *τ* 2 is scalar and used for all *P* .

## *Combining Kernels using Multiplication*

As I show in Section 4, feature bandwidth is important in spectral clustering: changes to the feature bandwidth (*θ*) can seriously change solutions. Therefore, let’s make explicit the parameters of the feature kernel. Following the same strategy from Section 2.3, a combined kernel can include bandwidths for both feature and geographic kernels:

**L** = **D** *−* **A***f* (*θ*) *◦* **A***s*(*η*) (4)

Here, **D** is again defined as the appropriate degree matrix obtained from the row-sums of **A***f* (*θ*) *◦* **A***s*(*η*). This specification is called *spatially-encouraged spectral clustering* since, as discussed below, cluster solutions are *encouraged* (but not required) to obey spatial constraints.

## *Combining Kernels Using Addition*

Despite superficial similarities, the multiplicative kernel combination discussed in Sec- tion 2.3 is a useful departure from previous attempts at introducing geography into spectral clustering. Most of these efforts derive from Wang and Davidson (2010)’s anal- ysis of constrained spectral clustering where the affinity matrix is a weighted average of spatial constraints and feature affinity:

**L***s* = **D***s −* [(1 *− δ*)**A***f* + *δ***A***s*] (5)

The new parameter, *δ*, controls the trade-off between spatial constraints and feature homogeneity. As *δ* becomes large, the constraints become increasingly important to the solution, while when *δ* = 0, the clustering only considers feature affinity.

## *Advantages of Multiplicative Kernel Combination*

Ultimately, the additive form behaves quite differently from the multiplicative form suggested in Eq. 4. First, the additive method means that clustering can only be “constrained” (*δ* = 1) or “relaxed” to some extent (*δ <* 1). When “constrained,” it does not consider feature information at all. Further, in additive combination, two observations (*i* and *j*) can be linked if their affinity is large enough (*Aij >* 0), even if they are not near one another ([*As*]*ij* = 0) or are a “must not link” pair ([*As*]*ij* = *−*1). While *δ* influences when this trade-off takes place, it also depends on the feature and kernel bandwidths. In multiplicative combination, constraints are always preserved.

Second, the “relaxation” of constraints can be done in an explicitly geographical fashion in the multiplicative method, but this is not necessarily done in the additive method. For example, in a simple binary spatial kernel, where 1 is used to mean observations can be linked and 0 means they cannot be linked, increasing *η* will increase the topological or geographical distance at which observations can be connected. This is an explicitly *geographical* relaxation of constraints. Changes to *δ* in the additive strategy provide no similar guarantees.

Third, the resulting affinity matrix tends to be much more sparse. In kernel combi- nation with multiplication, the final matrix is only as sparse as the sparsest component. In contrast, an additive kernel combination is only as sparse as the *least* sparse compo- nent. Practically, this is important: **A***s* is usually orders of magnitude more sparse than **A***f* , and obtaining the top *K* eigenvectors from a sparse matrix (while not effortless)

is much easier than obtaining them for a dense matrix. Therefore, in big data, spectral clustering on the multiplication-combined kernel can be used where the feature (or additive) kernel may be computationally prohibitive.5 In general, this means kernel combination using multiplication is a better strategy when representing regionaliza- tion problems for spectral clustering: the constraints you specify will be obeyed and the problem will be more tractable.

# Interrogating Spatially-Encouraged Spectral Clustering

In addition to being at the frontier of image segmentation and unsupervised learning methods, spatially-encouraged spectral clustering is useful for a variety of geograph- ical clustering problems. However, its novelty means that its semantics are relatively poorly understood. Below, I explain the meaning of a few structures in this model of spatial clustering. First, the interaction between the feature and spatial kernels is shown in Section 4.1 with an example clustering electoral data in Texas counties Sec- ond, new findings are reconciled with Yuan et al. (2015)’s results in Section 4.2. The new findings suggest that spectral clustering can provide a strong basis for a flexi- ble clustering technique for geographic data science. Finally, an additional example detecting competition clusters in Brooklyn Airbnb data is shown.

## *Solutions Depend on Both Geographical and Feature Similarity*

Both feature and spatial bandwidth parameters matter when combining kernels using multiplication. In unconstrained spectral clustering, changes in the feature kernel can give dramatically different solutions (Von Luxburg 2007, p.18). For constrained spec- tral clustering, it is the *combined* geographic-feature kernel that governs the solution. Any change to the combined kernel can significantly change the results. Surprisingly, this happens regardless of whether those changes arise from feature or spatial band- width changes, as shown below.

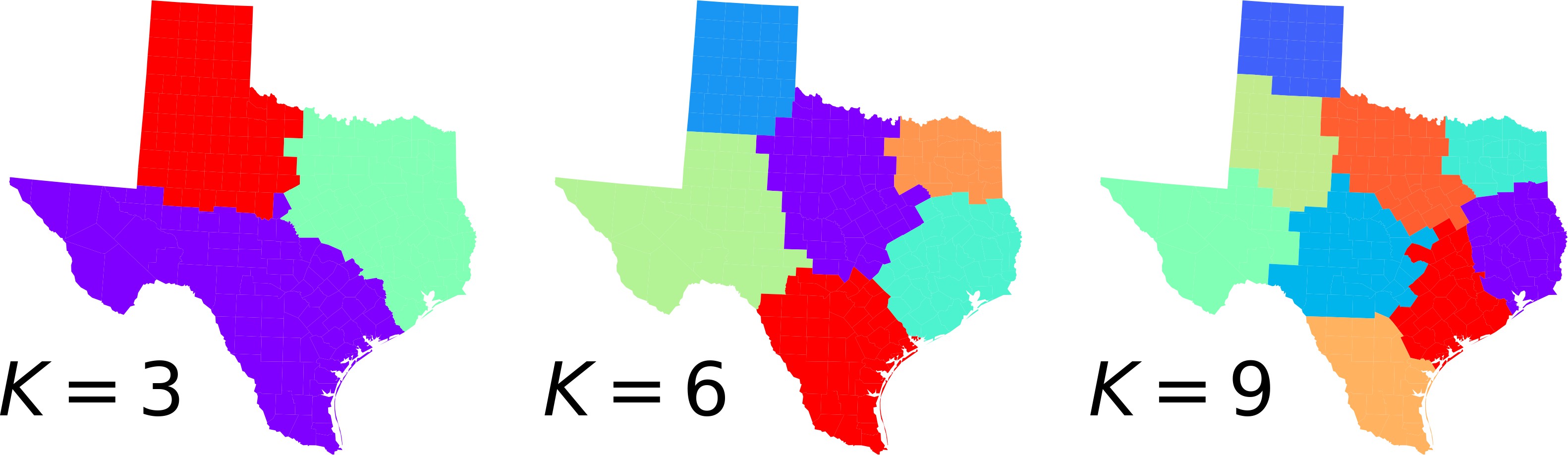
*A purely geographical example*

To start, it helps to examine one extreme for spatially-encouraged spectral clustering: when only geographic information is considered. In this case, spectral clustering is a community detection algorithm on a the graph arising from the geographical struc- ture of our data. Only geographical position/relation are considered, and the features observed at each site are ignored.

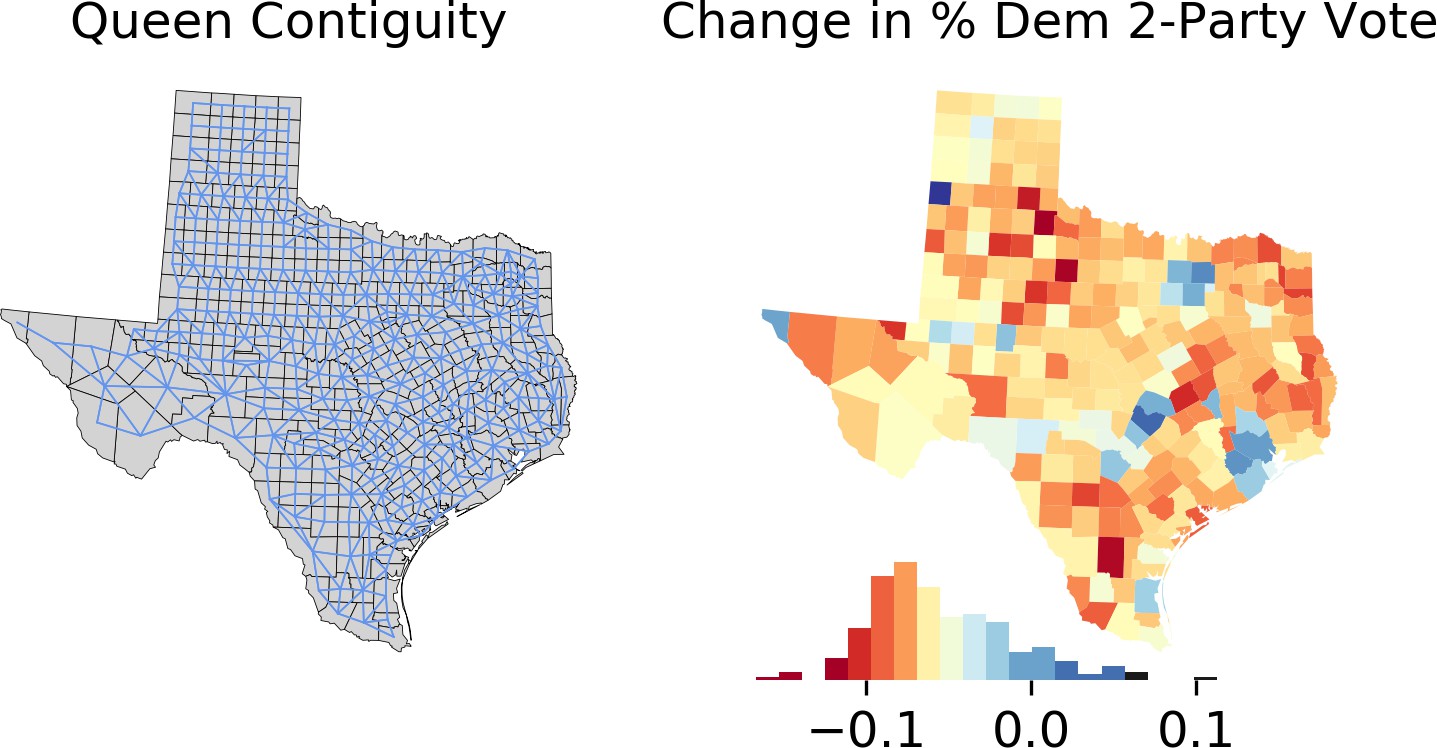
For this, let us use the linear contiguity kernel suggested by Yuan et al. (2015)—a first-order contiguity/adjacency matrix **A***s*0. Then, let all feature affinities be 1 so that the combined kernel remains unchanged, **A***f ◦* **A***s*0 = **A***s*0. Spectral clustering approxi- mates a normalized minimum weighted cut problem on this binary affinity matrix (Shi and Malik 2000). So, the solutions will tend towards geographically compact regions of similar size. Whereas this spatial regularity is an explicit secondary objective in other approaches (Li et al. 2014), this tends to emerge by default from spectral clustering.

An example of this is shown for Texas counties in Figure 1, finding three, six, and nine clusters. Note the clusters are fairly evenly-sized and quite compact. This case,

5This algorithm is made available post-publication in PySAL, the Python spatial analysis library (Rey and Anselin 2007), and is built primarily using NumPy (van der Walt et al. 2011) and scikit-learn (Pedregosa et al. 2011).



**Figure 1.** Clusters in the graph spectrum of queen contiguity in Texan counties.



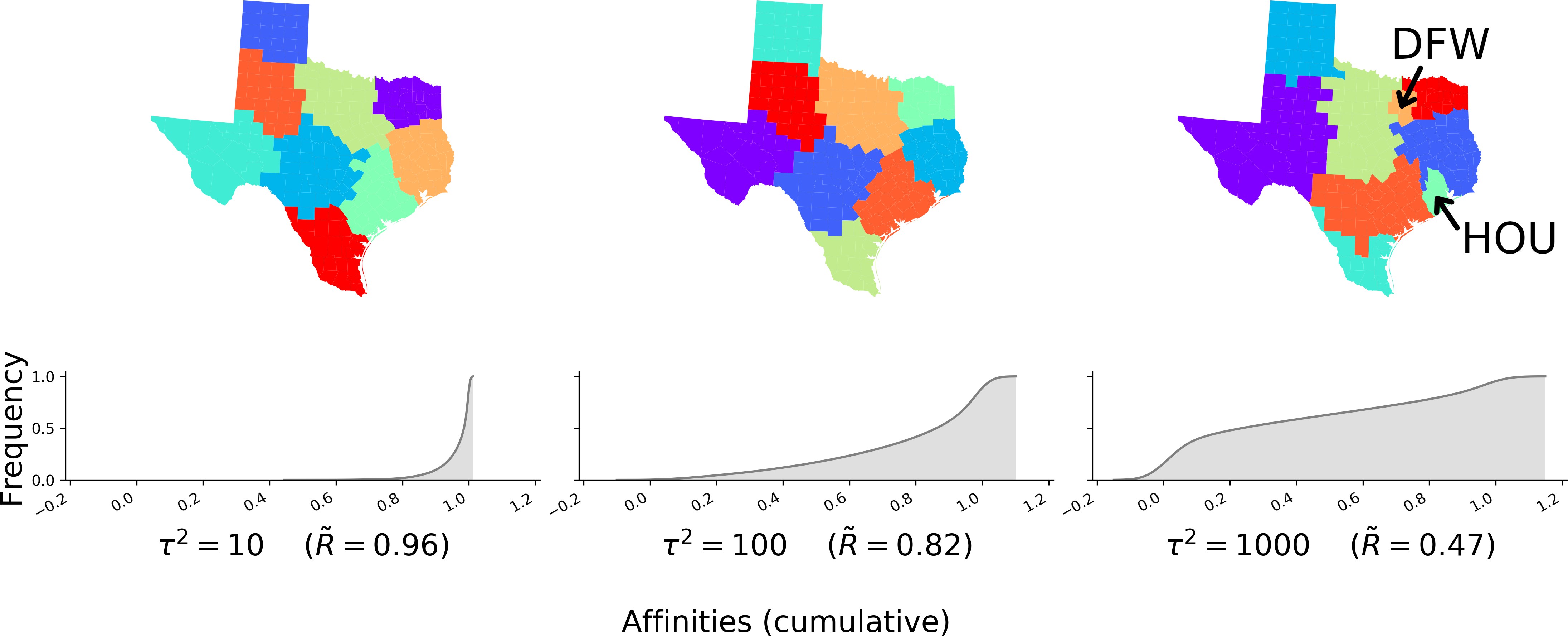
**Figure 2.** Adjacency graph and change in two-party vote share in Texas, 2008 to 2016. The chromatic legend is centered around the average change in Democrat two-party vote share (*−.*06). Negative red values indicate that the district swung more towards Republicans than Texas as a whole from 2008 to 2016.

where only spatial structure is considered, represents one end of the trade-offs between latent spatial structure in **A***s*0 and latent feature geography in **A***f* . The other end of the trade-offs will depend on the feature clusters that arise when all observations are considered equally “connectible,” and **A***s* is a matrix of ones.

*Introducing the Feature Kernel*

To proceed, let’s introduce the feature kernel and examine how changes in *τ* 2 begin to deviate from the purely-spatial solutions. For simplicity, consider a single feature: the change in the two-party vote for US president from 2008 to 2016 (Figure 2). This map shows several regions of the state where counties tended to swing together in the same direction. This means their vote share *intensified* towards one party, even if that party did not win the county.

In theory, the purely-spatial solution will dominate when the feature kernel values are constant, and when the feature kernel values vary significantly, then both spatial and feature information will result in new solutions. In practice, Figure 3 shows this behavior using a map of clusters on the top row and the distribution of feature affinities corresponding to the clusters on the bottom row. The *R*˜ in parenthesis is the adjusted



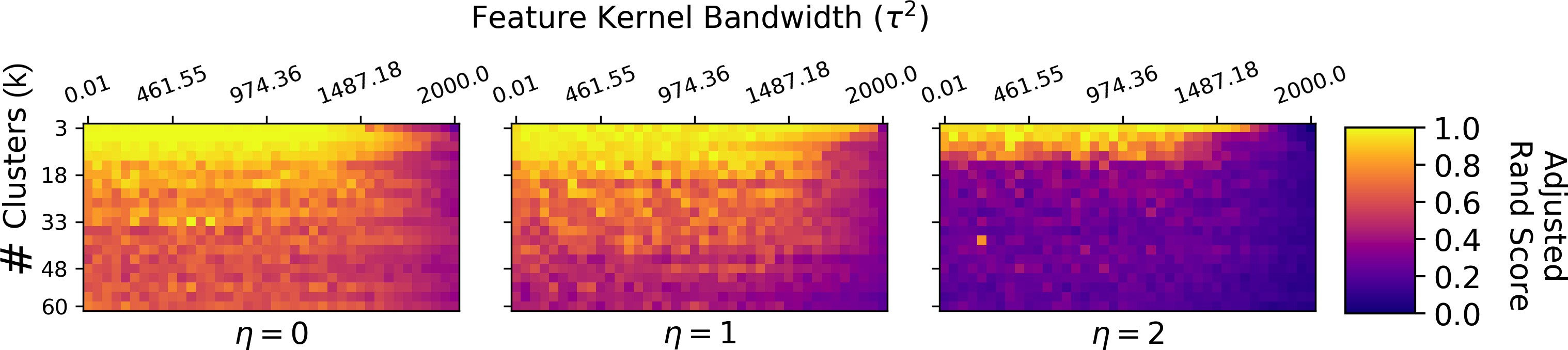
**Figure 3.** Spectral clusters and affinities for the presidential swing in Texas counties. Moving right, *τ* 2 increases. In all cases, the “linear spatial kernel” is used, so *η* = 0 always. The top row reflects the ultimate solution to the clustering problem for *k* = 9 clusters, and the bottom row reflects the cumulative distribution of feature affinities contained in **A***f* for the specified *τ* 2.

Rand score (Hubert et al. 1985), that describes the similarity between the purely- geographic partition and the partition arising from our combined feature-geographic kernel. The Rand score varies between 0 and 1, with 0 indicating random/unassociated labeling and 1 indicating that the two sets of labels are the same. The maps in that figure are produced for *K* = 9, but *K* is illustrative here, and is independent of the affinities shown below each map; any *K* could be chosen for a given *τ* 2 value and a similar result visualized. Further, no adjustment to the spatial kernel has been made: in Yuan et al. (2015)’s terms, the linear spatial kernel (i.e. standard adjacency matrix, **A***s*0) is used throughout, so clusters are always strictly contiguous.

Figure 3 starts with a value of *τ* 2 that results in nearly all affinities tightly clustered around 1, and then proceeds to a value of *τ* 2 that evenly distributes affinities along (0*,* 1) to a final value of *τ* 2 where a plurality of observations have very small feature affinity. Put simply, from left to right, the affinity matrix moves from dense (mainly 1) to sparse (mainly 0). In these maps, there is a change in clusters due solely to the change in *τ* 2. For example, consider the emergence of the five-county clusters; one cluster centered on Dallas-Fort Worth (labeled “DFW”) and another centered Houston (labeled “HOU”) become visible in the last panel. These only emerge as the feature affinities become sufficiently varied (*τ* 2 *>≈* 700). The remaining solutions are largely dominated by spatial constraints. However, increasing *τ* 2 while *η* = 0 never breaks contiguity: all identified clusters remain internally connected, though they begin to become much more heterogeneous in size and are significantly distinct from the pure- geographic clusterings. This demonstrates that the feature kernel can affect the extent to which constraints are satisfied in spatial-spectral clustering—*η* is not the sole factor.

## *Understanding Spatially-Encouraged Spectral Clustering*

Given that the clustering solution may be sensitive to *τ* 2, it is important to identify whether the parameters *τ* 2 and *η* interact. Practically, this would mean that changes to the geographic bandwidth will have the same effect at any value of the feature bandwidth. This would also make tuning simple: set *τ* 2 to a reasonable “default” value and find the best *η*. Then, tune *τ* 2 while holding *η* constant. However, if the two



**Figure 4.** Similarity to the purely-spatial solution for varying the number of clusters (*k*), feature decay parameter (*τ* ), and spatial bandwidth (*η*) values in the Texas example. As *η >* 2, solutions become significantly less similar to the purely spatial solution for *k >* 6.

parameters interact, then there is no guaranteed “acceptable” default value of *τ* 2, so they must be optimized at the same time.

To show whether the parameters are independent, a solution grid is shown in Figure 4 for values of *τ* 2, and *k* as *η* is increased.6 The color in each cell shows the Rand score for comparing the solution obtained at that (*η, τ* 2*, k*) configuration to the “purely- geographic” spectral clustering solution obtained at (*η,* 0*, k*). That is, these panels become more yellow when the obtained solution is more like the purely-geographic clusters at order *η*; they are blue when the combined kernel gives clusters that are very distinct from the purely-geographic clusters.

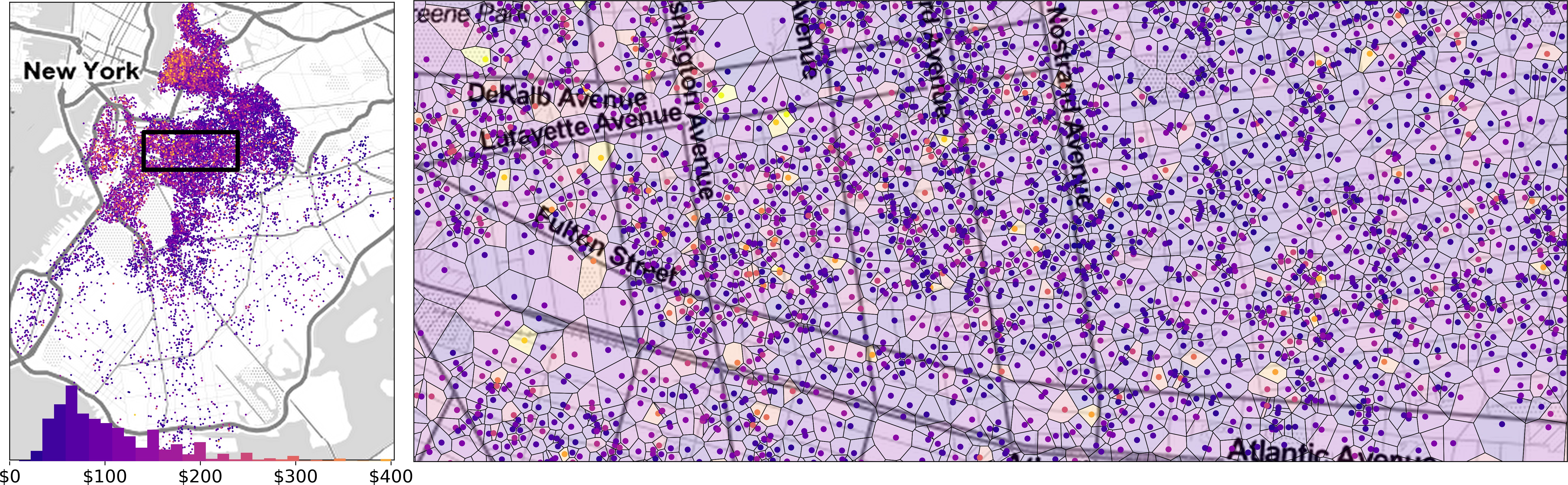
Solutions are shown to depend on both *η* and *τ* , and they do interact according to Figure 4. Moving within a single panel of Fig. 4, solutions are weakly sensitive to *τ* 2 when *η* = 0; the similarity with the purely-spatial solution declines as *τ* 2 increases. This effect becomes more dramatic when *η >* 0. Further, nearly all solutions where *η >* 2 (not shown in Figure 4) have very different clusterings from when *η* = 0, as is expected. Second, two values of *τ* 2 can differ strongly in their effects. When *τ* 2 = 2000, *η* = 2, and *k* = 3, the solution is has nearly zero similarity to that obtained from the geographical information alone. When *τ* 2 *<* 1500 with *η* = 2 and *k* = 3, however, solutions are strongly similar to the geography-alone clustering. Thus, this means two things are true: (1) *η*’s effects depend on *τ* 2 and (2) *τ* 2’s effects depend on *η*. Thus, spatial spectral clustering must be considered a two-parameter clustering algorithm, not a one-parameter clustering algorithm.

So, why do Yuan et al. (2015) suggest that when *η* increases, “constraint satisfaction becomes more relaxed” (p. 4)? For an arbitrarily-picked *τ* 2 value, this may be true, but it can change for other values of *τ* 2. In fact, as *η* changes, *τ* 2 can go from strongly affecting constraint relaxation to barely affecting constraint relaxation at all. While it is true that sparser affinity kernels (i.e. larger *τ* 2) will increase the heterogeneity in link costs and drive the solution to be more distinct from the purely-spatial solution, the degree to which this holds *for a specific η* is not known in advance. Therefore, it is necessary to tune both spatial and feature parameters.

# Spatially-encouraged Spectral Clustering for non-lattice data

Because the form of the spatial kernel is flexible, spatially-encouraged spectral cluster- ing can be used for data in any geographic support—point data, as well as for polygons.

6The binarized contiguity kernel is used here for simplicity. Each row of Fig. 4 uses the **A***η* connectivity matrix, connecting observations with maximum path order *η*, since the non-binary exponential kernel behaves substantively similarly to *η* = 0.



**Figure 5.** Prices of Airbnb rentals in Brooklyn, NY. The distribution of prices over all of Brooklyn (which is highly skewed) is also shown in the chromatic legend. The region of the city enclosed in the dashed box on left is shown in detail on the right, showing the Voronoi cells used for clustering.

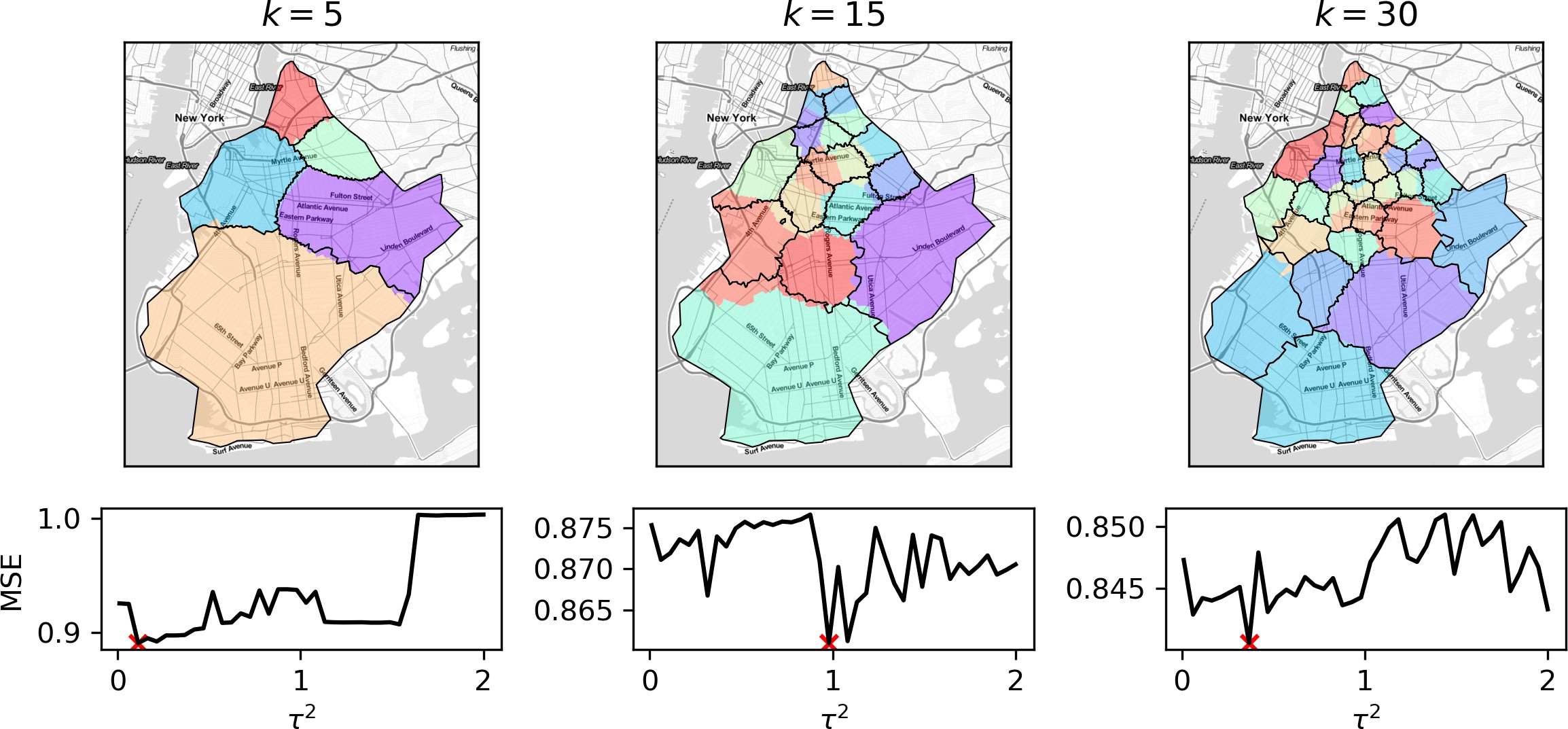
Point clustering is sometimes treated with formally-distinct methods because cluster- ing *locations* that arise from a continuous field is conceptually distinct from clustering regions from fixed “lattice” positions. However, spectral clustering works well in this instance as well. For point data, spatially-encouraged spectral clustering only requires a symmetric graph describing relations between points and an appropriate feature kernel.

To show that this approach works on a realistic dataset, I find price clusters in the prices of Airbnb listings in Brooklyn, point-referenced data. By comparison, many density-based techniques (such as DBSCAN (Ester et al. 1996) or Hierarchical- DBSCAN (McInnes et al. 2017)) are used in geography to identify clusters of spatial collocation without considering the features of the points themselves (e.g. Pavlis et al. 2018). However, spectral clustering can compute affinities for *features*, and then group locations that are both close and similar. This assurance of both geographical co- herence and feature homogeneity means that spatial spectral clustering can identify emergent geographic regions where prices tend to be similar to one another. Unlike other techniques, however, geographic coherence is directly parameterized by *η*, and the combination of kernels ensures that the geographic coherence is satisfied in the solution for some hyperparameter values, as discussed below.

## *Airbnb Price Clusters*

Airbnb, a popular ”sharing economy” service, offers an alternative to typical hotel accommodation. Notably, the geographical extent and intensity of its market penetra- tion is much more widespread in Brooklyn than the equivalent presence of hostels and hotels. Further, the intensity of market prevalence is highly varied within Brooklyn: some areas have very high density of listings and others have very low density. This strong difference in densities within a map is traditionally challenging for density-based clustering methods, but is not for spectral clustering. Using spectral clustering, larger price-homogeneous regions can be detected even as the city “stretches out” towards its periphery; clusters in listing price there will still reflect areas where prices are both similar and *relatively* close. Thus, these areas will reflect price convergence zones, or markets where Airbnbs have similar external amenities due to their extreme spatial proximity, so competition is effective only on price and listing amenities.

Identifying price clusters requires that price homogeneity and spatial coherence are balanced. Price clusters should represent a contiguous “formal area” where landlords



**Figure 6.** Pure spatial-spectral and spatially-encouraged spectral clustering maps for Brooklyn Airbnbs. The colored polygons indicate the pure spatial-spectral solutions, and the dark outlines represent the boundaries for the optimal spatially-encouraged solution considering listing price.

price their properties to compete with substitutes in that area. Practically speaking, this means a “price cluster” must balance geographic coherence and feature homo- geneity, so spatially-encouraged spectral clustering may be useful where other common point clustering approaches only focus on spatial coherence.

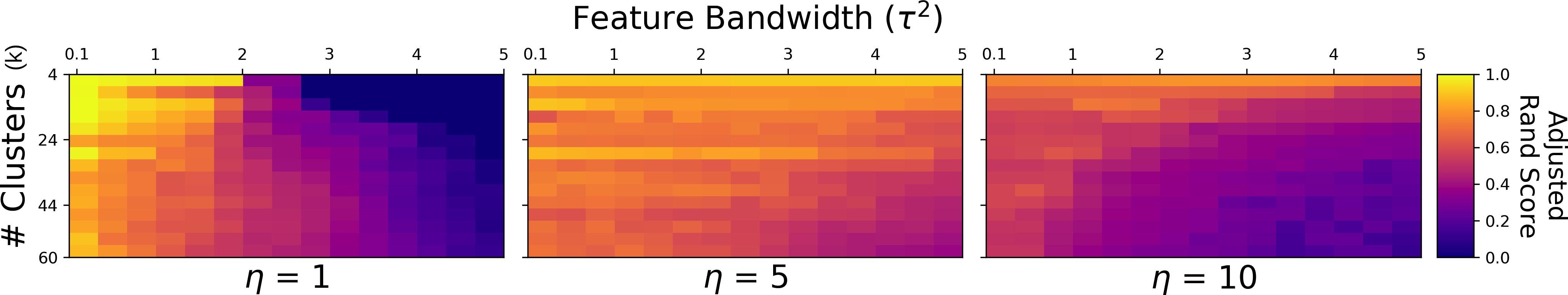
To apply spatially-encouraged spectral clustering on the log nightly price of Airbnb listings in Brooklyn, the Delaunay triangulation was constructed. This provides an adjacency graph that can be used as a spatial affinity matrix, **A***s*0, which structures the “geographic” component of the clustering.7 The feature kernel function, **A***f* , is a Gaussian kernel as before, and structures the “feature” component of the clustering. A detailed map of this price data is shown in Figure 5, alongside a macro-level image of the structure of Airbnb prices both in distribution and across Brooklyn.

Only relying on the geographical structure encoded by the Delaunay triangulation, the clusters in Figure 6 show clusters in the Brooklyn Airbnbs when price is either ignored or included for three different cluster counts. In all scenarios, strict spatial contiguity constraints are used, so *η* = 0, but the feature parameter *τ* 2 is allowed to vary. The colored polygons represent the “purely spatial” solution, and the black boundaries represent the solution obtained by an optimal combined kernel solution. In the lower row of the plot, the out-of-sample mean squared error for each *τ* 2 value is shown, with a red marker denoting the best mean squared prediction error for a 25% hold out sample.8 This demonstrates that adding feature information makes very little difference to cluster fit or the boundaries of clusters when *k* is small. But, for the *k* = 15 and *k* = 30 cases shown above, introducing feature information substantially changes cluster boundaries. The mean squared error for those solutions improves slightly, and all encouraged solutions consistently have better MSE.

Second, as before, feature and spatial kernel bandwidths must be optimized jointly

7Further, this has similar semantics to the Queen contiguity matrix used in the previous example: the Delaunay triangulation is the dual graph of a Voronoi diagram for the Airbnbs, as the Queen contiguity graph is a kind of dual graph for the Texas counties. Their order statistics, both at first order and higher, are also similar. Alternative spatial kernels, like *k*-nearest neighbor or Distance-weighted kernels could also be used.

8Precisely, I set aside 25% of the listings, compute their nearest geographic cluster, and predict their price using the mean cluster price. I am grateful to an anonymous reviewer for proposing this method.



**Figure 7.** Label concordance between the purely-geographic (*τ* 2 = 0) and spatially-encouraged solutions for values of *τ* 2 and *η*.

if spatial constraints are relaxed. Figure 7 shows the relationship between the feature bandwidth, spatial bandwidth, and number of clusters in the style of Figure 4 from the Texas voting example. Here, too, we see similar behavior. When *τ* 2 or the number of clusters (*k*) is small, solutions tend to be similar to the purely-spatial solution, but this changes depending on *η*. This suggests that the two parameters interact in this example, too, so it is unlikely they can be tuned independently as previously assumed in the literature.

# Discussion

In total, spectral clustering and kernel combination methods hold significant promise for geographic clustering, regionalization, and exploratory spatial data analysis. First, this paper shows that using multiplication to combine spatial and feature kernels in an affinity-based clustering algorithm has distinct advantages. Instead of additive combinations that require an extra parameter, multiplicative kernel combination can be tuned directly with the bandwidths of geographic and feature kernels. Multiplicative methods also make it clearer when a constraint, such as geographic contiguity, will be satisfied or not. And, since many efficient affinity-based clustering methods exist, this kernel combination strategy could be used beyond spectral clustering.

Second, this paper demonstrates the effectiveness of multiplicative kernel combi- nation using spectral clustering. Regionalization is already recognized as a computa- tionally hard problem (Duque et al. 2007), and spectral clustering makes it easy to generate “good” regions since it tends to discover contiguous clusters with roughly even size (Shi and Malik 2000). This has been explicitly modeled in other integer pro- gramming explorations (Li et al. 2014), so pure spatial-spectral clustering should be explored further in that domain.

Finally, the pairing of spectral clustering with a kernel combination, called ”spatially-encouraged spectral clustering” in this paper, can blend regionalization and typologization tasks. Using a “wide” geographic kernel and a “narrow” feature kernel will provide geodemographic-style *map typologization* that finds the most homoge- neous classes possible. Using a narrow spatial kernel and a wide feature kernel will provide solutions focus more on geographic coherence in the identified regions, result- ing in compact and evenly-sized clusters. Using these directly allows analysts to make a precise trade-off between feature and geographic coherence.

Conceptually, this method is very flexible. Geographical kernels in spatially- encouraged spectral clustering need only to describe symmetric geographical relations, and so can be used on a wide variety of spatial support. It can be highly local first-order contiguity among polygons or lines, it may model point proximity within a given dis- tance, or may itself be a continuous Mat´ern covariance function over a field. Likewise,

feature kernels can represent distances that are relevant for the data at hand. The feature affinity structure could be a straightforward feature kernel on continuous data (e.g. Gaussian, bisquare, or triangular kernels), or represent string similarity useful in neighborhood sequence analysis (Delmelle et al. 2013, e.g.), or even could represent compositional similarity for a set of sub-populations (Chodrow 2017). Optimizing their combination can be done using grid search (as employed here for illustration) or more advanced optimization methods, such as simulated annealing or genetic algorithms. Since the parameters governing this trade-off can be tuned, the measurement of “co- herence”may also be extend to a more general idea of contiguity (Wu and Murray 2008) or new hybrid measures of spatial-social homogeneity (Wolf et al. 2019).

# Conclusion

This paper identifies an opportunity to extend and expand (Yuan et al. 2015) and takes it. By examining an omitted parameter in their discussion, I suggest a technique for *spatially-encouraged spectral clustering* that is significantly more general. Further, this more general method offers the end user more control over the structure of the solution, since both spatial and feature kernels are able to be tuned. This makes the trade-offs explored by Yuan et al. (2015) significantly sharper, and allows us to identify exactly the endpoints of these trade-offs, the purely data-driven and geography-driven partitions.

As such, the specification of spatially-encouraged spectral clustering is a fundamen- tal advance in the adaptation of spectral clustering methods to geographical prob- lems. While it may be undesirable to introduce additional tuning parameters, this paper makes these parameters explicit only because they existed implicitly anyway. By making them explicit and demonstrating that they interact, we show that it is necessary to consider both spatial and feature kernel structure when drawing regions. Further, work that ignores these free parameters may fail to generalize in unexpected ways, as changes in the structure of a spatial kernel have different impacts depend- ing on the structure of the feature kernel. As such, this work is a novel, important, and generally more useful technique for cluster analysis in geographic data science. Spatially-encouraged spectral clustering provides a framework that can blend *typol- ogization*, used in geodemographics, and *regionalization*, used in spatial optimization and statistics. Together, this provides a novel explicit specification that represents how geographical coherence and cluster fit compete for relevance in unsupervised learning problems.

# Data and Codes Availability Statement

All code and documentation for the plots and algorithms in this paper are made avail- able on the Open Science Framework (https://doi.org/10.17605/OSF.IO/FCS5X). Further, a generalized *spatially-encouraged spectral clustering* algorithm has been made available in the PySAL package (Rey and Anselin 2007) as part of the spopt subpack- age. The algorithm depends primarily on NumPy (van der Walt et al. 2011) and scikit-learn (Pedregosa et al. 2011).

# Additional Information

## *Funding*

This material is based upon work supported by the National Science Foundation under Grant No. 1733705, as well as the Alan Turing Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation and/or the Alan Turing Institute.

# References

Jared Aldstadt and Arthur Getis. Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis*, 38(4):327–343, October 2006. ISSN 0016- 7363, 1538-4632. .

Luc Anselin. Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2):93–115, 1995.

Giuseppe Arbia, Giuseppe Espa, and Danny Quah. A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics*, 34:81–103, 2008.

J. Besag and J. Newell. The dection of clusters in rare diseases. *Journal of the Royal Statistical Society A*, 154(1):143–155, 1991.

Philip S Chodrow. Structure and information in spatial segregation. *Proceedings of the Na- tional Academy of Sciences*, 44:11591–11596, 2017.

S Czamanski and L Ablas. Indentification of industrial clusters and complexes: A comparison of methods and findings. *Urban Stud.*, 16:61–80, 1979.

Nema Dean, Guanpeng Dong, Aneta Piekut, and Gwilym Pryce. Frontiers in Residential Segregation: Understanding Neighbourhood Boundaries and Their Impacts. *Tijdschrift voor economische en sociale geografie*, in print, 2018. ISSN 1467-9663. .

Elizabeth Delmelle, Jean-Claude Thill, Owen Furuseth, and Thomas Ludden. Trajectories of Multidimensional Neighbourhood Quality of Life Change. *Urban Studies*, 50(5):923–941, 2013. .

Guanpeng Dong, Levi John Wolf, Alekos Alexiou, and Dani Arribas-Bel. Inferring neighbour- hood quality with property transaction records by using a locally adaptive spatial multi-level model. *Computers, Environment and Urban Systems*, September 2018. ISSN 0198-9715. .

Marjan Drukker, Charles Kaplan, Frans Feron, and Jim van Os. Children’s health-related quality of life, neighbourhood socio-economic deprivation and social capital. A contextual analysis. *Social Science & Medicine*, 57(5):825–841, September 2003.

Juan C Duque, Raul Ramos, and Jordi Surinach. Supervised Regionalization Methods: A Survey. *International Regional Science Review*, 30(3):195, 2007.

Juan C Duque, Richard L Church, and Richard S Middleton. The p-Regions Problem. *Geogr.*

*Anal.*, 43(1):104–126, 2011.

Juan C Duque, Luc Anselin, and Sergio J Rey. The max-p-regions problem. *Journal of Regional Science*, 52(3):397–419, 2012.

Martin Ester, Hans-Peter Kriegel, J¨org Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

G. Galster. On the nature of neighbourhood. *Urban studies*, 38(12):2111, 2001.

Arthur Getis and J. Keith Ord. Local spatial statistics: An overview. *Spatial Analysis: Mod- elling in a GIS Environment*, 374, 1996.

Daniel A Griffith. Eigenfunction properties and approximations of selected incidence matrices employed in spatial anlaysis. *LInear Algebra and its Applications*, 321:95–112, 2000.

Daniel A Griffith. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Springer Science & Business Media, 2013.

Tony H Grubesic, Ran Wei, and Alan T Murray. Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense. *Annals of the Association of American Geographers*, 104(6):1134–1156, November 2014. ISSN 0004-5608, 1467-8306. .

Richard Harris, Peter Sleight, and Richard Webber. *Geodemographics, GIS and Neighbourhood Targeting*, volume 7. John Wiley and Sons, 2005.

Richard Harris, Ron Johnston, and Simon Burgess. Neighborhoods, Ethnicity and School Choice: Developing a Statistical Framework for Geodemographic Analysis. *Population Re- search and Policy Review*, 26(5):553–579, 2007. ISSN 1573-7829. .

Trevor J. Hastie, Robert John Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-84857-0.

Lawrence J Hubert, Reginald G Golledge, C Michael Costanzo, and Nathan Gale. Measur- ing association between spatially defined variables: An alternative procedure. *Geographical Analysis*, 17(1):36–46, 1985.

W Isard. Regional science, the concept of region, and regional structure. *Pap. Reg. Sci.*, 2(1): 13–26, 1956.

Geoff M. Jacquez, Andy Kaufmann, and Pierre Goovaerts. Boundaries, links and clusters: A new paradigm in spatial analysis? *Environmental and Ecological Statistics*, 15(4):403–419, December 2008. ISSN 1352-8505, 1573-3009. .

Kamyoung Kim, Yongwan Chun, and Hyun Kim. P-Functional Clusters Location Problem for Detecting Spatial Clusters with Covering Approach. *Geographical Analysis*, 49(1):101–121, 2017. ISSN 1538-4632. .

Elijah Knaap, Levi John Wolf, Sergio Rey, and Wei Kang. The Dynamics of Urban Neighbor- hoods: A Survey of Approaches for Modeling Socio-Spatial Structure. *Geography Compass*, Under Review, 2019. .

M. Kulldorff and N. Nagarwalla. Spatial disease clusters: Detection and inferenc. *Statistics in Medicine*, 14:799–810, 1995.

Wenwen Li, Richard L Church, and Michael F Goodchild. The p-compact-regions problem.

*Geographical Analysis*, 46(3):250–273, 2014.

Leland McInnes, John Healy, and Steve Astels. Hdbscan: Hierarchical density based clustering.

*The Journal of Open Source Software*, 2(11):205, 2017.

Alan T Murray, Tony H Grubesic, and Ran Wei. Spatially significant cluster detection. *Spatial Statistics*, 10:103–116, 2014.

D.B. Neill, A.W. Moore, F. Pereira, and T. Mitchell. Detecting significant multidimensional spatial clusters. *Advances in Neural Information Processing Systems*, 17:969–976, 2005.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.

Johannes Oehrlein and Jan-Henrik Haunert. A cutting-plane method for contiguity- constrained spatial aggregation. *Journal of Spatial Information Science*, (15), December 2017. ISSN 1948-660X. .

Michalis Pavlis, Les Dolega, and Alex Singleton. A Modified DBSCAN Clustering Method to Estimate Retail Center Extent: Clustering Retail Agglomerations. *Geographical Analysis*, 50(2):141–161, April 2018. ISSN 00167363. .

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blon- del, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Jour- nal of Machine Learning Research*, 12:2825–2830, 2011.

Sergio J Rey and Luc Anselin. PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*, 37(1):5–27, 2007.

Sergio J. Rey and Daniel J. Mattheis. *Identifying Regional Industrial Clusters in California: Volume I Conceptual Design*. 2000.

P Rogerson and I Yamada. *Statistical Detection and Surveillance of Geographic Clusters*.

Chapman & Hall/CRC, 2009.

Taylor Shelton and Ate Poorthuis. The Nature of Neighborhoods: Using Big Data to Rethink the Geographies of Atlanta’s Neighborhood Planning Unit System. *Annals of the American Association of Geographers*, 109(5):1341–1361, September 2019. ISSN 2469-4452, 2469-4460.

.

Jianbo Shi and Jitendra Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), August 2000.

Alex Singleton and Paul A Longley. Creating open source geodemographics: Refining a national classification of census output areas for applications in higher education. *Pap. Reg. Sci.*, 88 (3):643–666, August 2009.

Alex Singleton and Seth E Spielman. The Past, Present, and Future of Geodemographic Research in the United States and United Kingdom. *Prof. Geogr.*, 66(4):558–567, 2014.

Seth E Spielman and John R Logan. Using high-resolution population data to identify neigh- borhoods and establish their boundaries. *Annals of the Association of American Geogra- phers*, 103(1):67–84, 2013.

B. W. Turnbull, E. J. Iwano, W. S. Burnett, H. J. Howe, and L. C. Clark. Monitoring clusters of disease: Application to Leukemia incidence in upstate New York. *American Journal of Epidemiology*, 132(1):136–143, 1990.

St´efan van der Walt, S. Chris Colbert, and Ga¨el Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13(2):22–30, March 2011. ISSN 1521-9615. .

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Xiang Wang and Ian Davidson. Flexible constrained spectral clustering. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 563–572. ACM, 2010.

Xiang Wang, Buyue Qian, and Ian Davidson. On constrained spectral clustering and its applications. *Data Mining and Knowledge Discovery*, 28(1):1–30, January 2014. ISSN 1384- 5810, 1573-756X. .

Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 274–285. SIAM, 2005.

Levi John Wolf, Elijah Knaap, and Sergio Rey. Geosilhouettes: Geographical measures of cluster fit. *Environment and Planning B: Urban Analytics and City Science*, page 2399808319875752, September 2019. ISSN 2399-8083. .

X. Wu and Alan T Murray. A new approach to quantifying spatial contiguity using graph theory and spatial interaction. *International Journal of Geographical Information Science*, 22(4):387–407, April 2008. ISSN 1365-8816, 1362-3087. .

Shuai Yuan, Pang-Ning Tan, Kendra Spence Cheruvelil, Sarah M. Collins, and Patricia A. Sorrano. Constrained spectral clustering for regionalization: Exploring the trade-off between spatial contiguity and landscape homogeneity. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference On*, pages 1–10. IEEE, 2015.

Di Zhu, Fan Zhang, Shengyin Wang, Yaoli Wang, Ximeng Cheng, Zhou Huang, and Yu Liu. Understanding Place Characteristics in Geographic Contexts through Graph Convolutional Neural Networks. *Annals of the American Association of Geographers*, 110(2):408–420, March 2020. ISSN 2469-4452. .

Merle Zwiers, Reinout Kleinhans, and Maarten Van Ham. The Path-Dependency of Low- Income Neighbourhood Trajectories: An Approach for Analysing Neighbourhood Change. *Appl. Spatial Analysis*, pages 1–18, 25 2016.